

# Micro-level Estimation of Child Malnutrition Indicators and Its Application in Cambodia

Tomoki Fujii\*

JEL classification code: C15, I12, I32, O15

Key words: Cambodia, Concentration curve, Decomposition, Health inequality, Malnutrition, Small-area estimation, Targeting

World Bank Policy Research Working Paper 3662, July 2005

*The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the view of the World Bank, its Executive Directors, or the countries they represent. Policy Research Working Papers are available online at <http://econ.worldbank.org>.*

\*Singapore Management University and the World Bank. email: [fujii@are.berkeley.edu](mailto:fujii@are.berkeley.edu). I am deeply indebted to Chris Elbers, Jean Olson Lanjouw and Peter Lanjouw for their advice from the beginning of this research. I also thank Alain de Janvry, Livia Montana, Mahadevan Ramachandran, Martin Ravallion, H. E. Kim Saysamalen, Elisabeth Sadoulet, Boreak Sik and H.E. San Sythan. I also thank the Government of Japan under the Millenium PHRD grant for financial support. Of course, all the remaining mistakes are mine.

# 1 Introduction

Malnutrition remains a major public health concern in most developing countries. The World Health Organization (2002) estimates that about 3.7 million deaths among young children worldwide were related to malnutrition in 2000. Similarly, Pelletier et al. (1994) estimate that about one-half of childhood deaths in developing countries are caused by undernutrition. Malnutrition has also been associated with mortality and morbidity in later life, delayed mental development, decreased cognitive and behavioral functioning throughout childhood and adolescence and poorer performance in school (de Onis et al., 2000; Galler and Barrett, 2001; Glewwe et al., 2001; Shariff et al., 2000).

In Cambodia, almost half of the children under five are malnourished as measured by the height-for-age or weight-for-age indicator (National Institute of Statistics et al., 2001). Given the grave consequences of malnutrition, the issue deserves serious attention. However, as with many other developing countries, the resources available for improving children's nutritional status are severely limited in Cambodia. Hence, efforts must be made to allocate the available resources in an efficient manner. In particular, geographic targeting, or targeting according to locational information, is often easy to administer and implement, and can be quite effective when malnourished children are concentrated in certain locations. However, the formulation of an effective geographic targeting policy requires knowledge of the location of malnourished children, information which is often not readily available. The Cambodia Demographic and Health Survey (CDHS) 2000 data provide information on child nutritional status at the level of 17 strata, a level that is more aggregated than the provincial level.<sup>1</sup> Hence, this information is too aggregated to be useful for formulating targeting policies.

Such informational constraints are among the central issues concerning the formulation of targeting policies (Ravallion and Chao, 1989; Kanbur, 1987). This study aims to overcome

---

<sup>1</sup>In Cambodia, there are four administrative divisions, which are province, district, commune and village in descending order of aggregation. On average, they have about 89,000, 12,000, 1,300 and 160 households respectively. Each stratum has on average 125,000 households.

the problem by producing commune-level estimates of the prevalence of malnutrition in Cambodia. The estimates can be projected onto maps, which allow policy-makers to visually identify areas of severe child malnutrition, analyze the current situation of malnutrition and formulate geographic targeting policies aimed at assisting the neediest people in a more efficient and transparent manner.

To derive the commune-level estimates, we combine the CDHS 2000 dataset with individual level Cambodian Population Census data for 1998. The former includes information on child nutrition status but has a limited number of observations, while the latter covers virtually everyone in Cambodia but lacks any specific information on child nutrition status. The approach we take builds on the small-area estimation technique developed by Elbers, Lanjouw and Lanjouw (2000; 2002; 2003a, hereafter ELL). We extend their methodology to jointly estimate multiple indicators and allow for a richer structure of error terms, a critical step to address issues unique to nutrition indicators.

While the commune-level estimates of the prevalence of malnutrition are in themselves of interest for policy-makers, we take the analysis one step further and illustrate three distinct but related applications of the methodology. First, we investigate the relationship between consumption poverty, inequality and health at the commune level. While we find no simple relationship, we find that non-linear effects of consumption on the nutritional status of children are important for understanding the relationship. Second, we decompose health inequality indicators into between group and within group components by geographic information. This application is useful for elucidating the significance of the geographic information in explaining overall inequality. We propose two decomposable inequality indicators that are useful for this purpose. One is based on the analysis of variance. The other uses the concentration curve, which is similar to the Lorenz curve but orders individuals on the horizontal axis according to the group they belong to instead of the individual ranking. Concentration curves have been used to analyze the health inequality across different socioeconomic groups and

the effect of taxation on income distribution (*e.g.* Kakwani (1977); Kakwani et al. (1997); Mahalanobis (1960); Wagstaff and van Doorslaer (2004); Yitzhaki and Slemrod (1991)).

Third, we evaluate the potential gains from geographic targeting in the presence of commune-level estimates. We show that the concentration curve is useful for this purpose as well. We assume that we can target resources only at the stratum level with the DHS alone. Given a fixed budget, we can compare reductions in child malnutrition with stratum-level and commune-level targeting. Alternatively, we can calculate the reduction in resource needs to achieve the same goal if targeting is carried out at a smaller geographic level. In general, the efficiency gain depends on the spatial distribution of malnourished children. We find that the savings in the cost of nutrition programs from commune-level targeting is on average at least two to three times higher than that from stratum-level targeting when the *per capita* cost of the program is fixed. The efficiency gains can be much higher under alternative assumptions.

This paper is structured as follows: Section 2 reviews the measurement and prediction of nutritional status of children. Section 3 develops the methodology of nutrition mapping. We then discuss the data in Section 4, followed by the results in Section 5. Section 6 discusses the applications of the methodology, and Section 7 concludes.

## **2 Measurement and prediction of child nutrition status**

To measure malnutrition in a non-invasive and inexpensive manner, anthropometry has been widely used among nutritionists and epidemiologists. Among the most commonly used anthropometric measures are weight-for-height, weight-for-age and height-for-age Z-scores. Z-scores measure the number of standard deviations between an individual's value of the anthropometric indicator and the median of the National Center for Health Statistics (NCHS) growth reference population of the same sex, and in the same age or height group. Deficiency in weight-for-height, weight-for-age and height-for-age Z-scores is respectively called "wast-

ing”, “underweight”, and “stunting.” We use the conventional cut-off point of negative two to calculate the prevalence of malnutrition. For example, the prevalence of stunting is defined as the number of children with a height-for-age Z-score below negative two over the total number of children. (See Dibley et al. (1987a,b); Waterlow et al. (1977); WHO Working Group (1986, 1995) for further discussion on Z-scores.)

As WHO Working Group (1986) points out, there are several obvious differences among these measures. First, one can lose weight but not height. Second, linear growth is a slower process than growth in body mass. Third, catch-up in height is possible, but takes a relatively long time even with a favorable environment. Thus, wasting reflects ‘acute’, or short-term, malnutrition whereas stunting reflects ‘chronic’, or long-term, malnutrition with underweight somewhere in between. Given these differences, it should not be surprising if patterns of wasting and stunting are different. In fact, Victora (1992) finds no systematic pattern that holds for an international population between levels of stunting and wasting.

As with Pradhan et al. (2003), we use standardized height and weight, which are the Z-scores converted back to the corresponding height and weight of the reference age-sex group of 24-month-old girls. The standardized height and weight are an affine transformation of Z-scores and preserve all the desirable properties that the original Z-scores possess. The additional merit of the standardized height and weight is that they are always positive for practically possible values of Z-scores, and that we can compute inequality measures in terms of height or weight. We chose the reference group of 24-month-old girls in order to make the inequality decomposition analysis in this study comparable to the world inequality decomposition by Pradhan et al. (2003).

The standardized height and weight corresponding to a negative two Z-score is analogous to the “poverty line”  $z$ , a cut-off point below which an individual is considered poor. We can, therefore, define the FGT measure (Foster, Greer and Thorbecke, 1984) of malnutrition by simply replacing the consumption by the standardized height or weight. Hence, letting  $y_i^{(1)}$

and  $y_i^{(2)}$  be individual  $i$ 's standardized height and weight, and  $z^{(1)}$  and  $z^{(2)}$  be the standardized height and weight corresponding to negative two Z-score, the FGT measure of malnutrition with parameter  $\theta$  can be written as follows:

$$P^{\theta,(k)} = \sum_i \text{Ind}(y_i^{(k)} < z^{(k)}) \cdot \left( \frac{z^{(k)} - y_i^{(k)}}{z^{(k)}} \right)^\theta \quad (k = 1, 2)$$

where  $\text{Ind}(\cdot)$  denotes the indicator function.  $P^{0,(1)}$  and  $P^{0,(2)}$  are the prevalence of stunting and underweight respectively.

The methodology we develop below is built on the association between anthropometric indicators and other socio-economic and geographic indicators. It is, therefore, instructive to briefly overview the previous studies on the relationship between anthropometric indicators and other indicators.<sup>2</sup> We report on the prevalence of stunting and underweight, but not wasting. The primary reason for this is that we were unable to construct a regression model for weight-for-height with sufficient explanatory power.<sup>3</sup> Interestingly, this seems to be the case in other countries. Alderman (2000) creates regression models of various anthropometric indicators for Vietnam, South Africa, Pakistan and Morocco, and the variability of weight-for-height is least well-captured among all the anthropometric indicators. His regression results suggest that community level effects are of great importance. We also found that the commune-level variables are important for explaining the variation of anthropometric indicators.

Li et al. (1999) investigate the issue of malnutrition with various anthropometric indices and examine its correlates in a large sample of poor rural minority children in China. In this study, age, maternal height, water sources, maternal education and very low income were significant correlates. In a study for Vietnam, similar factors were found to be relevant. Height-for-age Z-scores were significantly correlated with the age of the child, maternal weight

---

<sup>2</sup>Note also that our model is a predictive one and not intended to describe causal relationship.

<sup>3</sup>It is possible to estimate weight-for-height from weight-for-age, height-for-age and age. This is a straightforward extension of this research.

and height, parental education, and some indicator variables on access to water sources (Haughton and Haughton, 1997). In West Africa, residence in a dry zone was found to be associated with wasting, but not with stunting once other variables are introduced as controls (Curtis and Hossain, 1998). Frongillo et al. (1997) make a cross-country comparison and find that higher energy availability, female literacy and gross domestic product were the most important factors associated with lower prevalence of stunting.

Monteiro et al. (1997) investigate the patterns of intra-familial distributions of under-nutrition in Brazil. They analyze data for four income strata separately and find that under-nutrition is significantly associated among household members for the 25 percent poorest families. Khorshed Alam Mozumder et al. (2000) look into the effects of the length of the interval between children on malnutrition in two districts in Bangladesh. They conclude that the results indicate the potential importance of longer birth intervals in reducing child malnutrition. Zeini and Casterline (2002) explore the importance of different levels of geographic clustering in Egypt, including regional level, governorate-level, local level, household level and individual level, using the 2000 Egypt Demographic and Health Survey. They find that spatial clustering does seem to be important in that country. They also find that, even after controlling for socioeconomic factors, significant household-level clustering remained and, interestingly, that individual clustering (*i.e.* across indicators) was also important. Furthermore, they find that underweight was correlated with stunting and/or wasting in some, but not all, governorates of Egypt.

### 3 Methodology

**Overview** We describe in this section the methodology used to estimate the indicators of malnutrition at the level of small geographic areas. The methodology developed here is similar to the small area estimation procedure developed by ELL (2000; 2002; 2003a) for

estimating consumption poverty and inequality in that we also combine survey data with unit record census data in order to obtain estimates at a lower level of aggregation than the survey permits. Their small area estimation approach was first applied to Ecuador (Hentschel et al., 2000) and has subsequently been applied to poverty and inequality measures in many countries, including Cambodia (Alderman et al., 2002; Demombynes et al., 2002; Fujii, 2004)). We first provide a general overview of the methodology, and then describe it more formally.

The basic idea is straightforward. We first construct a prediction model of anthropometric indicators using only the variables that are common between the census and the survey, along with geographic indicators available for the entire country at the village or commune level. The geographic indicators include the remotely-sensed data as well as village-level statistics derived from the census data. Common geographic codes in all data sets allow these to be linked to both the survey and the census data.

The parameters of the model are estimated with the survey dataset. An important feature of our study and of the ELL approach is the explicit treatment of the error terms. We estimate regression coefficients and the associated variance-covariance matrix, and also scrutinize the distribution of the disturbance terms in order to carry out simulation. In each round of simulation, we randomly draw regression coefficients and disturbance terms in accordance with their estimated distribution and we impute the anthropometric indicators to each census record. By geographically aggregating the imputed anthropometric indicators, we can estimate the prevalence of malnutrition.

There are four major differences between this study and earlier work by ELL. The first and most obvious difference concerns the type of the survey dataset used for estimation. ELL have focused on consumption or income taken from a socio-economic survey, while we have anthropometric measures taken from a Demographic and Health Survey (DHS). A second difference stems from different units of analysis. Consumption data are usually produced at the household level, whereas anthropometric measures are at the individual level. In the



ELL approach, disturbance terms are decomposed into a location-specific effect (usually at the cluster-level) and a household-specific effect. However, in our study, it is important to allow as well for an unobserved individual-specific effect in addition to location- and household-specific effects.

The third difference is related to the second point. The number of children under five in a household is very limited, with no more than two for most of the households. Hence, we cannot rely on large-sample properties when estimating the parameters or distribution of the individual-specific effect. Thus, unlike the ELL approach, we make finite-sample corrections. The fourth point is the number of left-hand-side variables used in this study. In the ELL approach, only a single consumption or income measure is considered, but we consider multiple indicators. Because the unobserved parts of the different indicators may be correlated, this must be taken into account when computing the parameter estimates. We call correlation across indicators the intra-personal effect, and allow for it in the model. Given that weight-for-age is a medium-term indicator of nutritional status, which is partly affected by height-for-age, it would be natural to assume that intra-personal effect may exist. Consequently, our approach involves simultaneous estimation of several models, unlike the single-model procedure applied by ELL (2002; 2003a).

**Parameter Estimation** We denote the set of all clusters by  $\mathcal{C}$ , the set of all households in cluster  $c(\in \mathcal{C})$  by  $\mathcal{H}_c$  and the set of all the children under five in household  $h(\in \mathcal{H}_c)$  by  $\mathcal{I}_{ch}$ . Let  $y_{chi}^{(k)}$  be the  $k$ -th ( $1 \leq k \leq K$ ) anthropometric indicator of interest for the individual  $i$  in cluster  $c$  and household  $h$ .<sup>4</sup> In our particular application,  $K = 2$ , with  $k = 1$  and  $k = 2$  being the standardized height and weight respectively, but in general  $K$  could be larger. Our goal is to find an estimate of an aggregate index  $\mathbf{W}_{\mathcal{V}}^{(k)} \equiv \mathbf{W}(\{y_i^{(k)}\}_{i \in \mathcal{V}})$  such as the prevalence of malnutrition for a set of individuals  $\mathcal{V}$ .  $y_{chi}^{(k)}$  is related to a  $d^{(k)}$ -vector of

---

<sup>4</sup>A cluster is the primary sampling unit for the survey, which is a village for CDHS 2000.

observable characteristics,  $\mathbf{x}_{chi}^{(k)}$ , through the following anthropometric model.

$$y_{chi}^{(k)} = \left[ \mathbf{x}_{chi}^{(k)} \right]^T \beta^{(k)} + u_{chi}^{(k)}$$

where  $\beta^{(k)}$  is a  $d^{(k)}$ -vector of parameters and  $u_{chi}^{(k)}$  is a disturbance term.  $u_{chi}^{(k)}$  satisfies  $E[u_{chi}^{(k)} | \mathbf{x}_{chi}^{(k)}] = 0$  for all  $c, h, i$  and  $k$ . Let  $C \equiv \#(\mathcal{C})$ ,  $H_c \equiv \#(\mathcal{H}_c)$ , and  $I_{ch} \equiv \#(\mathcal{I}_{ch})$ , where  $\#(\cdot)$  is the counting measure. The total number of observations is  $N \equiv \sum_{c \in \mathcal{C}} \sum_{h \in \mathcal{H}_c} I_{ch}$  and each cluster has a weight  $w_c$  which is normalized so that  $\sum_c w_c = 1$ . Recall here that the number of children under five in one household is usually small and can be 1 for some households.

As with the ELL study, we allow for location, or cluster-specific, effect and heteroskedastic household-specific effect. In addition, we have multiple indicators and the individual effect correlated across the indicators. Hence,  $u_{chi}^{(k)} = \eta_c^{(k)} + \epsilon_{ch}^{(k)} + \delta_{chi}^{(k)}$ , where  $\eta_c^{(k)}$ ,  $\epsilon_{ch}^{(k)}$ , and  $\delta_{chi}^{(k)}$  are respectively the location, household and individual effect. In principle, each of the three components of  $u_{chi}^{(k)}$  could be heteroskedastic and correlated across the indicators. The particular choice we make is driven by the insights from previous studies and also by the limitation imposed by the data. For example, given that  $\mathbf{I}_{ch} = 1$  for many households in the CDHS data, it is extremely difficult to distinguish household effects from individual effects if both are heteroskedastic. Cluster-level heteroskedasticity is also difficult to estimate because of the limited number of clusters. We allow for flexibility in the correlational structure of the disturbance terms where such flexibility seems most crucial.

We shall hereafter denote  $(u^{(1)}, \dots, u^{(K)})^T$  by  $\vec{u}$ , and use similar notation for other variables. We assume that  $\vec{\eta}_c, \vec{\epsilon}_{ch}$  and  $\vec{\delta}_{chi}$  are uncorrelated and satisfy  $E[\vec{\eta}_c] = E[\vec{\epsilon}_{ch}] = E[\vec{\delta}_{chi}] = \vec{0}_K$ , where the last term is the  $K$ -vector of zeros. For  $l, k \in \{1, 2, \dots, K\}$  with  $l \neq k$ , we assume  $E[\eta_c^{(k)} \eta_c^{(l)}] = E[\epsilon_{ch}^{(k)} \epsilon_{ch}^{(l)}] = 0$ . We denote the variance of each component of the disturbance term as  $(\sigma_\eta^{(k)})^2 \equiv E[\eta_c^{(k)} \cdot \eta_c^{(k)}]$ ,  $(\sigma_{\epsilon, ch}^{(k)})^2 \equiv E[\epsilon_{ch}^{(k)} \cdot \epsilon_{ch}^{(k)}]$ , and  $(\sigma_\delta^{(k)})^2 \equiv E[\delta_{chi}^{(k)} \cdot \delta_{chi}^{(k)}]$ . Note that we need subscripts  $ch$  to express the heteroskedasticity of the household effect.

We also assume that the intra-personal effect is  $\sigma_\delta^{(l,k)} = \sigma_\delta^{(k,l)} \equiv E[\delta_{chi}^{(l)} \cdot \delta_{chi}^{(k)}]$  so that the intra-personal correlation is  $\rho^{(l,k)} = \frac{\sigma_\delta^{(l,k)}}{\sigma_\delta^{(k)} \sigma_\delta^{(l)}}$ . We shall denote simple means by dots such as  $u_{ch.}^{(k)} = \frac{1}{I_{ch}} \sum_{i \in \mathcal{I}_{ch}} u_{chi}^{(k)}$  and  $u_{c..}^{(k)} = \frac{1}{H_c} \sum_{h \in \mathcal{H}_c} u_{ch.}^{(k)}$ .

As discussed above, we estimate the distribution parameters of the estimate of the regression coefficients and disturbance term. To do so, we run an ordinary least squares (OLS) regression for indicator  $k$ , and get the residual  $\hat{u}_{chi}^{(k)}$ . Letting  $\tilde{\mathcal{H}}_c \equiv \{h \in \mathcal{H}_c | I_{ch} > 1\}$ ,  $\tilde{H}_c \equiv \#\{\tilde{\mathcal{H}}_c\}$ ,  $\tilde{\mathcal{C}} \equiv \{c \in \mathcal{C} | \tilde{H}_c > 0\}$ , and  $\tilde{w}_c \equiv \frac{w_c}{\sum_{c' \in \tilde{\mathcal{C}}} w_{c'}}$ , straightforward calculations give

$$E \left[ \sum_{c \in \tilde{\mathcal{C}}} \frac{\tilde{w}_c}{\tilde{H}_c} \sum_{h \in \tilde{\mathcal{H}}_c} \sum_{i \in \mathcal{I}_{ch}} \frac{(u_{chi}^{(k)} - u_{ch.}^{(k)})^2}{I_{ch} - 1} \right] = (\sigma_\delta^{(k)})^2$$

and

$$E \left[ \sum_{c \in \tilde{\mathcal{C}}} \frac{\tilde{w}_c}{\tilde{H}_c} \sum_{h \in \tilde{\mathcal{H}}_c} \sum_{i \in \mathcal{I}_{ch}} \frac{(u_{chi}^{(k)} - u_{ch.}^{(k)}) \cdot (u_{chi}^{(l)} - u_{ch.}^{(l)})}{I_{ch} - 1} \right] = \sigma_\delta^{(k,l)}.$$

We obtain consistent estimators  $(\hat{\sigma}_\delta^{(k)})^2$  and  $\hat{\sigma}_\delta^{(k,l)}$  by taking out the expectations operator on the left-hand-side and replacing  $u$  by  $\hat{u}$  in the respective equation. In the same manner, we obtain a consistent estimator  $(\hat{\sigma}_\eta^{(k)})^2$  of the variance of location effect using the following formula (proof in the Appendix).

$$E \left[ \frac{\sum_{c \in \mathcal{C}} w_c H_c (u_{c..}^{(k)})^2 - \sum_{c \in \mathcal{C}} \frac{w_c}{H_c} \sum_{h \in \mathcal{H}_c} (u_{ch.}^{(k)})^2}{\sum_{c \in \mathcal{C}} w_c (H_c - 1)} \right] = (\sigma_\eta^{(k)})^2 \quad (1)$$

There is no guarantee that the estimated variance of the location effect is non-negative, and hence we censor  $(\hat{\sigma}_\eta^{(k)})^2$  at zero.

The household and individual effects are difficult to separate from each other because  $I_{ch}$  is small for a majority of households. To estimate the distribution parameters of the household effect, it is useful to work with the sum  $(s_{ch}^{(k)})^2 (\equiv (\sigma_{\epsilon, ch}^{(k)})^2 + (\sigma_\delta^{(k)})^2)$  of the household and individual effects. Note that the heteroskedasticity of  $(s_{ch}^{(k)})^2$  comes only from the household effect. We can show the following formula (proof in Appendix) for  $c \in \mathcal{C}^* \equiv \{\psi \in \mathcal{C} | H_\psi > 2\}$ :

$$E \left[ \frac{H_c \cdot (u_{ch}^{(k)} - u_{c\cdot}^{(k)})^2}{H_c - 2} - \frac{\sum_{h' \in \mathcal{H}_c} (u_{ch'}^{(k)} - u_{c\cdot}^{(k)})^2}{(H_c - 1)(H_c - 2)} \right] + \frac{I_{ch} - 1}{I_{ch}} (\sigma_\delta^{(k)})^2 = (s_{ch}^{(k)})^2 \quad (2)$$

We let  $\hat{s}_{ch}^2$  be the left hand side of the equation above with the expectation operator removed and with  $u$  and  $\sigma_\delta$  replaced by  $\hat{u}$  and  $\hat{\sigma}_\delta$  respectively.<sup>5</sup> For the heteroskedastic model, we propose the following logistic heteroskedastic model similar to the one in ELL (2003a):

$$\ln \frac{(\hat{s}_{ch}^{(k)})^2 - B^{(k)}}{A^{(k)} + B^{(k)} - (\hat{s}_{ch}^{(k)})^2} = [\mathbf{z}_{ch}^{(k)}]^T \alpha^{(k)} + \tau_{ch}^{(k)}$$

where  $A^{(k)}$  and  $B^{(k)}$  are the maximum and minimum of  $(s_{ch}^{(k)})^2$ , and  $\mathbf{z}_{ch}^{(k)}$  is vector of household characteristics,  $\alpha^{(k)}$  heteroskedastic regression coefficient and  $\tau_{ch}^{(k)}$  the residual term. The feature of this formulation is that  $(\hat{s}^{(k)})^2$  is both upper- and lower-bounded. Using  $B_*^{(k)} = \min\{0, 1.05 \cdot \min_{ch}\{(\hat{s}_{ch}^{(k)})^2\}\}$  and  $A_*^{(k)} = 1.05 \cdot (\max_{ch}\{(\hat{s}_{ch}^{(k)})^2\} - B_*^{(k)})$ , we run OLS to obtain  $\hat{\alpha}^{(k)}$ . The use of the delta method suggests the following estimate of  $(\sigma_{\epsilon, ch}^{(k)})^2$ :

$$(\hat{\sigma}_{\epsilon, ch}^{(k)})^2 = \max \left\{ 0, \left[ \frac{D^{(k)}}{1 + D^{(k)}} + \frac{D^{(k)}(1 - D^{(k)})(\hat{\sigma}_\tau^{(k)})^2}{2(1 + D^{(k)})^3} \right] A_*^{(k)} + B_*^{(k)} - (\hat{\sigma}_\delta^{(k)})^2 \right\} \quad (3)$$

where  $D^{(k)} \equiv \exp([\mathbf{z}_{ch}^{(k)}]^T \hat{\alpha}^{(k)})$ , and  $(\hat{\sigma}_\tau^{(k)})^2$  is the estimated variance of  $\tau_{ch}$ . The  $\max\{\cdot, \cdot\}$  function is introduced to ensure the non-negativity of  $(\hat{\sigma}_{\epsilon, ch}^{(k)})^2$ . The consequence of this is that  $(\sigma_{\epsilon, ch}^{(k)})^2$  may be upward-biased. Hence, the standard errors for the estimates of nutrition measures at the level of small geographic areas are conservative. We can now estimate the variance-covariance matrix  $\Omega$ , and carry out a (feasible) generalized least squares (GLS) regression to obtain the regression coefficients  $\hat{\beta}$  and  $\widehat{Var}[\hat{\beta}]$  for all the indicators at once.

---

<sup>5</sup>When  $(\hat{\sigma}_\tau^{(k)})^2 = 0$ , we use  $E[(u_{ch}^{(k)})^2] = (\sigma_{\epsilon, ch}^{(k)})^2 + \frac{(\sigma_\delta^{(k)})^2}{I_{ch}}$  instead.

We then find the empirical distributions of each disturbance component. We approximate the distributions of  $\eta_c$ ,  $\epsilon_{ch}$ , and  $\delta_{chi}$  respectively by the distributions of  $u_{c\cdot}$ ,  $(u_{ch\cdot} - u_{c\cdot})$ , and  $(u_{chi} - u_{ch\cdot})$  standardized to have mean zero and a unit standard error.

**Simulation** We explicitly take into account the model and idiosyncratic errors using Monte-Carlo simulation, where the model error is the error associated with the estimation of model parameters  $\alpha$  and  $\beta$ , and the idiosyncratic error is the variations of the anthropometric indicators that remain unexplained by the model. Now, let  $R$  be the number of simulations, which must be sufficiently large in order to make the computational errors small enough. In our study, we set  $R = 100$ . In  $r$ -th simulation where  $r \in \{1, 2, \dots, R\}$ , we need the following parameters:  $\tilde{\beta}_{(r)}^{(k)}$ ,  $\tilde{\alpha}_{(r)}^{(k)}$ ,  $(\tilde{\sigma}_{\tau,(r)})^2$ ,  $(\tilde{\sigma}_{\delta,(r)}^{(k)})^2$ ,  $(\tilde{\sigma}_{\delta,(r)}^{(k,l)})^2$ ,  $(\tilde{\sigma}_{\eta,(r)}^{(k)})^2$ ,  $A_{*,(r)}^{(k)}$ ,  $B_{*,(r)}^{(k)}$  for  $\forall k \neq l$ . We randomly draw  $\tilde{\alpha}_{(r)}$  and  $\tilde{\beta}_{(r)}$  from the normal distribution with mean  $\hat{\alpha}$  and  $\hat{\beta}$ , and variance-covariance matrix  $\widehat{Var}[\hat{\alpha}]$  and  $\widehat{Var}[\hat{\beta}]$  respectively. For the rest of the parameters, we create a two-stage bootstrapping sample of  $\hat{u}$  in each round of simulation, and compute the parameters using the bootstrapping sample.<sup>6</sup> It is straightforward to calculate  $(\tilde{\sigma}_{\epsilon, ch, (r)}^{(k)})^2$  from  $(\tilde{\sigma}_{\tau, (r)})^2$ ,  $A_{*, (r)}^{(k)}$ ,  $B_{*, (r)}^{(k)}$  from Eq(3).

For each census record, we draw the standardized cluster, household and individual effects. Each of these disturbance components is drawn jointly for multiple indicators to capture the intra-personal effect. Letting the standardized components of disturbance terms drawn in the  $r$ -th simulation be  $\tilde{\eta}_{c, (r)}^{(k)}$ ,  $\tilde{\epsilon}_{ch, (r)}^{(k)}$ , and  $\tilde{\delta}_{chi}^{(k)}$  respectively, the  $k$ -th imputed anthropometric indicator for the individual in the census in  $r$ -th simulation is<sup>7</sup>:

$$\tilde{y}_{chi, (r)}^{(k)} = \mathbf{x}_{chi}^{(k)} \tilde{\beta}_{(r)}^{(k)} + \tilde{\eta}_{c, (r)}^{(k)} \cdot \tilde{\sigma}_{\eta, (r)} + \tilde{\epsilon}_{ch, (r)}^{(k)} \cdot \tilde{\sigma}_{\epsilon, ch, (r)} + \tilde{\delta}_{chi}^{(k)} \cdot \tilde{\sigma}_{\delta, (r)}$$

---

<sup>6</sup>Another possible implementation is to draw  $\tilde{\alpha}$  and  $\tilde{\beta}$  from the bootstrapping sample.

<sup>7</sup>To eliminate extreme values, we drop census observations for which the point estimate  $\mathbf{x}_{chi}^{(k)} \hat{\beta}^{(k)}$  is not in the range of  $y^{(k)}$  for at least one indicator. In each simulation, we censored  $\tilde{y}_{chi, (r)}^{(k)}$  at the minimum and maximum observed in the survey. We also censored  $D^{(k)}$  at the minimum and maximum.

This allows us to calculate the aggregate indicator of interest as  $\tilde{\mathbf{W}}_{(r),\mathcal{V}}^{(k)} = \mathbf{W}(\{\tilde{y}_{i,(r)}^{(k)}\}_{i \in \mathcal{V}})$  in  $r$ -th simulation. The point estimate and its associated standard error are estimated as the mean and standard deviation of  $\tilde{\mathbf{W}}_{(r)}^{(k)}$  taken over the simulation. For example, the point estimate of the prevalence of stunting for commune  $\kappa$  is as follows:

$$\hat{\mathbf{P}}_{\kappa}^{0,(1)} = \frac{1}{R \cdot \sum_{c \in \mathcal{C}_{\kappa}} \sum_{h \in \mathcal{H}_c} \mathbf{I}_{ch}} \cdot \sum_{r=1}^R \sum_{c \in \mathcal{C}_{\kappa}} \sum_{h \in \mathcal{H}_c} \sum_{i \in \mathcal{I}_{ch}} \text{Ind}(\tilde{y}_{chi,(r)}^{(1)} < z^{(1)})$$

## 4 Data

Our basic building blocks comprise a survey dataset, a census dataset and a dataset of geographic variables. The survey dataset we used is CDHS 2000, which was designed to collect health and demographic information for the Cambodian population with a particular focus on women of childbearing age and young children. The sample covered 12,236 households in 17 strata across the country. Data collection took place over a six-month period between February and July in 2000. In addition to detailed information about each household, its members, and housing characteristics, one-quarter of these households were systematically selected to participate in the anthropometric data collection. All children under 60 months of age in the sub-sampled households were weighed and measured. After excluding children for which information on height or weight is missing or implausible, 3,596 observations were used for this analysis (for further details, see National Institute of Statistics et al. (2001)). We first derived the height-for-age and weight-for-age Z-score measures, which were then converted to the height and weight for 24 month-old females with the same Z-score measures.

The second source of data is the Cambodian National Population Census, the first population census to be conducted in Cambodia since 1962. The census covered virtually all persons staying in Cambodia at the reference time of midnight of March 3, 1998.<sup>8</sup> The census data

---

<sup>8</sup>Due to military operations, about 0.5 percent of the population was not covered.

contain information on housing characteristics as well as information on each usual household member and visitors present on the reference night, including the relationship to the head of household, sex, age, marital status, migration, literacy, education and employment. The census also contained questions on fertility of females aged 15 and over, and infant mortality. The census dataset contains about 1.4 million records of children under five.

A set of geographic indicators is also used in this analysis. Because Cambodia has a rich collection of geographic data, indicators on a range of characteristics could be generated. These indicators included distance calculations, land use and land cover information, climate indicators, vegetation, agricultural production and flooding. A number of datasets from various sources were compiled into a GIS and these indicators were generated for all villages and communes in Cambodia. Very coarse resolution data were summarized at the commune level, while high resolution data were attributed to individual villages. Distances from villages to roads, other towns, health facilities, and major rivers were calculated from the center of the villages. Indicators based on satellite data with varying temporal resolutions included land use within the commune (agricultural, urban, forested, and so forth), the Normalized Difference Vegetation Index (NDVI) to proxy agricultural productivity, and the degree to which the area was lit by nighttime lights as a proxy of urbanization. Relatively stable indicators including soil quality, elevation, and various 30-year average climate variables were derived from other composite datasets. We have also generated the village-level means from the census data. It should be noted that the village-level means do not have to be taken from the variables that also exist in the CDHS dataset. This is because the village-level means, as with other geographic variables, can be linked to both the census and the survey datasets. Inclusion of these geographic variables and their cross terms with other individual-level and household-level variables has improved substantially the ability to explain the variation of anthropometric indicators.

## 5 Results

We constructed an anthropometric model in each of the five zones (“ecozones”) of Urban, Plain, Tonlesap, Coastal and Plateau. We combined provinces that are similar in agro-climatic and socio-cultural characteristics because some of the strata had too few observations to carry out meaningful analysis.<sup>9</sup> We ran regressions in each ecozone separately. While individual-level and household-level variables explain only 20 to 30 percent of the variation in the standardized height and weight, we were able to increase the explanatory power of the model to about 40 to 60 percent by including geographic variables and interaction terms. This is consistent with Curtis and Hossain (1998). We checked the robustness of the regression coefficients by randomly dropping some households or clusters as was done in ELL (2002).

The GLS regression results for the Coastal ecozone are presented in Table 1 and Table 2.<sup>10</sup> The point estimate of the variance of the location effect was zero in all strata, even though in some rounds of the simulation  $(\tilde{\sigma}_{\eta,(r)}^{(k)})^2$  was strictly positive because of bootstrapping. The average proportion of individual effect to the sum of all the disturbance components was found very high with the ratio of individual effect to the overall residuals  $\frac{(\hat{\sigma}_{\delta}^{(1)})^2}{(\hat{\sigma}_u^{(1)})^2}$  ranging from 0.81 to 0.99 and  $\frac{(\hat{\sigma}_{\delta}^{(2)})^2}{(\hat{\sigma}_u^{(2)})^2}$  ranging from 0.66 to 0.92. This in turn means that the household effect is relatively small, though it is in general not negligible. In all the ecozones, the magnitudes of intra-personal correlations as measured by  $\hat{\rho}^{(1,2)}$  were found high, ranging from 0.44 to 0.53. This strongly suggests the importance of the inclusion of intra-personal correlation.

After the predictions for standardized height and weight for each child in the census were made in each round of simulation, they were aggregated to the commune level in Cambodia to arrive at the prevalence of stunting and underweight. Because of missing data in the census and geographic datasets for a small number of communes, we obtained commune-level estimates for a total of 1,594 communes out of the 1,616 communes in Cambodia.

---

<sup>9</sup>We followed the definition of the ecozones previously used by WFP (2001).

<sup>10</sup>Other regression results are available from the author upon request.



To evaluate the reliability of the estimates, the ecozone-level CDHS prevalence of stunting and underweight were compared with the ecozone-level estimates in this study. Table 3 summarizes these results. The differences between the CDHS estimates and this study are within two standard errors of the CDHS estimates, suggesting that the differences can be attributed to the random errors. It should also be noted from Table 3 that the standard errors estimated with the CDHS alone are higher than those for this study except for the underweight model in Tonlesap. The standard errors from Coastal ecozone in the survey are quite high, and our methodology improved the ecozone level estimates quite substantially. We also arrive at the same conclusion when comparisons are made at the level of 17 strata.

We also checked the magnitude of standard errors associated with commune-level estimates of malnutrition prevalence. The simple means of the standard errors are 9.0 percent for stunting and 10.1 percent for underweight, while the simple means of the coefficient of variation are 3.8 percent for stunting and 4.2 percent for underweight. They are reasonably small as they are about the same magnitude as the standard errors in the survey at the ecozone level. We should note, however, that there are communes for which the standard errors are quite high, as the maximum standard error was 22.7 percent for stunting and 18.5 percent for underweight. Overall, the accuracy of commune-level estimates for stunting and underweight is about the same. It should be noted that, as the number of communes included in a nutrition program increases, the idiosyncratic component of the error tends to decrease. Hence, if a proposed nutrition intervention delivers assistance to a relatively large number of communes, high levels of standard errors for each commune are not necessarily worrisome.

Using a GIS, we converted the commune-level estimates into maps. Figures 1 and 2 are the maps of stunting and underweight at the commune level. These maps show the point estimates of the prevalence of malnutrition as of the census year 1998 at the commune level, and the orange and red areas represent bad situations.

As shown in Figure 3(a), stunting and underweight exhibit by and large similar patterns,

even though there is no simple geographic pattern of malnutrition. At the commune-level, the correlation between the estimates of prevalence of stunting and underweight is 0.33. Red areas are the communes where the prevalence of stunting and underweight is both over 45 percent, which is approximately the national average for both stunting and underweight. Green areas have on the other hand prevalence of stunting and underweight both less than 45 percent. In terms of the number of communes, the green and red communes account for about 65 percent of all the communes.

The prevalence of both stunting and underweight is high in the most densely populated parts of Cambodia surrounding Phnom Penh, the provinces of Kandal, Prey Veng, Svay Rieng, Kampong Cham, and Kampong Chhnang as shown red in Figure 3(b). While these provinces are geographically close to each other, the causes of malnutrition may be quite different. Inspection of possible causes elucidates this point. For example, the prevalence of diarrhea for children under five is 29.7 percent in Kandal whereas it is only 3.1 percent in Prey Veng. Similarly, the prevalence of fever, which is a primary manifestation of malaria and other acute infections in children, is 46.8 percent in Kandal and 4.0 percent in Prey Veng (National Institute of Statistics et al., 2001). On the other hand, the poverty rates in Kandal and Prey Veng are estimated at 18.4 percent and 53.1 percent (Fujii, 2004). This suggests that malnutrition in Prey Veng is likely to be driven mainly by poverty, or more specifically lack of caloric intake, whereas infectious diseases seem important causes of malnutrition in Kandal.

Figure 3(b) also shows some noticeable differences between stunting and underweight. For example, most areas of Kampong Speu shown in pink have high levels of stunting but not underweight. This suggests that the nutritional status has improved recently, which is perhaps partly because of the recent improvement in road access in this province. Low levels of stunting and high levels of underweight as shown orange in Figure 3 reflect recent aggravation of nutritional status. Possible causes of aggravation include increased incidence

of malaria and diarrhea and acute food shortage due to natural disaster.

While these maps are presented in a user friendly format, they do not take into account the fact that the commune-level estimates are subject to statistical errors. We can also present the maps in terms of the difference between the commune-level estimate and a reference level such as the national average divided by the standard error of the commune-level estimate. Figure 4(a) shows how significantly different the prevalence of stunting in each commune is from the national average (NA). The point estimate in red areas are more than two standard deviations higher than the NA. Orange areas have a point estimate between one and two standard deviations higher than the NA. Green and yellowish green areas are similarly defined. In yellow areas, the absolute value of the difference is less than one standard deviation. This representation would be useful to identify a small number of communes for a nutrition program.

Another useful way to present the results would be to depict the density of malnourished children as shown in Figure 4(b). This picture is completely different from Figure 1 because the population density is very high in Phnom Penh and surrounding provinces. This representation would be most relevant when a proposed policy intervention is likely to benefit the target location as a whole. Construction of health clinics might be an example of such a project. Health clinics are likely to help improve the nutrition status of children in areas like Kandal, where infectious disease is likely to be a cause of the problem and the distance to health facility is problem. In fact, 60.1 percent of women in Kandal reported that distance to health facility is a big problem in accessing health care for themselves. This contrasts with 12.9 percent in Prey Veng. Different representations shed light on different aspects of malnutrition, and most appropriate format would depend on the purpose and intended audience of the map.

The maps are in themselves useful, but the estimates we derived have a number of applications. We illustrate three applications of the estimates in the next section.

## 6 Applications

### 6.1 Consumption Poverty and Inequality as Correlates of Malnutrition

The first application we illustrate in this paper is the analysis of the relationship between consumption poverty, inequality and malnutrition. We utilize the estimates of consumption poverty and inequality derived by Fujii (2004) using the standard ELL approach. One way to look at the relationship is to overlay maps. Figure 5(a) and Figure 5(b) are the stunting and underweight maps overlaid with the poverty map. Table 6 gives the number of communes by poverty rate and prevalence of underweight. While there is not an obvious pattern between poverty and malnutrition, there are at least two notable features on these maps. First, the northern part of Siem Reap and southern part of Prey Veng have high levels of stunting, underweight and poverty. These places have food security problems because many people lack means of production there. Southern Prey Veng has been repeatedly hit by drought and flood. Second, the northeastern provinces of Kratie, Stung Treng and Ratanakiri have low levels of poverty and high levels of malnutrition. This may be partly because of diseases and poor child care in this region (See Hardy and Health Unlimited Ratanakiri Team (2001) for the case of Ratanakiri), both of which are not direct results of poverty. By overlaying nutrition map with other maps, we can investigate the plausible causes of malnutrition in different parts of the country.

To analyze the correlational structure of aggregate indicators, we first regressed estimates of malnutrition indicators on the estimates of the logarithm of the commune-level mean consumption  $\bar{Y}$  and the generalized entropy inequality measure with parameter 0 (GE(0)).<sup>11</sup> Table 4 shows the results.<sup>12</sup> Not surprisingly, we found that the logarithmic mean consump-

---

<sup>11</sup>GE(0) is also known as the Theil's L index. See Shorrocks (1980) for the discussion of generalized entropy measures.

<sup>12</sup>The standard errors in Table 4 are underestimated because they do not take into account the fact that

tion has a negative and significant coefficient. However, the significance goes away once provincial-level dummies are included. The coefficient on inequality ( $GE(0)$ ) is also negative and significant in both specifications. The negative relationship between the prevalence of malnutrition and inequality also holds for the Gini index. This is a surprising result because inequality has been associated with poor health outcomes in various geographic locations and at various level of aggregation (See, for example, Wilkinson (1996); Kawachi et al. (1997)), even though a majority of the studies use mortality as a health indicator and relatively few studies are done in developing countries.

What we found from this analysis has an important implication for the current debate on the significance of inequality on health outcomes. Some researchers in public health, including Wilkinson (1996), have argued that higher income inequality leads to worse health outcomes partly because income inequality leads to higher levels of psychosocial stress. Economists have been skeptical about this line of argument because of omission of factors other than income inequality (Deaton, 2003; Wagstaff and van Doorslaer, 2000). We also argue that the psychosocial explanation is unattractive because it fails to explain the negative correlation between the prevalence of malnutrition and poverty rate.

We can explain the negative correlation between the rate of malnutrition and consumption with Figure 6. The horizontal axis measures consumption and the vertical axis measures the probability density of consumption within the commune. Let us consider a situation where consumption has positive and non-linear effects on the status of child nutrition. For simplicity assume that a child is malnourished if and only if the consumption of the child falls short of a threshold  $X$ . Let the solid line A be the graph of the probability density function for the reference population. A mean-preserving spread would make the graph look like the dashed line B, and an increase in the mean consumption would correspond to the

---

$\bar{Y}$  and  $GE(0)$  are estimates, while the point estimate is likely to be unbiased ELL (2004a). The correction requires the correlation of model errors between poverty and nutrition models, which we do not know because they were estimated from different data sets. This point also applies to Table 5.

dotted line C. The prevalence of malnutrition for B is the smallest followed by C and A. This argument is consistent with Deaton (2003). It is also similar to the argument made by Ravallion (1988) and Haddad and Kanbur (1990), though they dealt with welfare risk and intra-familial distribution respectively.

To see if Figure 6 is reasonable, we also regressed average severity of malnutrition among malnourished children (*i.e.*  $\frac{P^2}{P^0}$ ) on  $\bar{Y}$  and GE(0) as well as provincial-level dummies. We found that the coefficient on GE(0) was neither significant for stunting nor underweight. Once we include the prevalence of malnutrition  $P^0$  on the right hand side, GE(0) has a positive and significant coefficient for both malnutrition indicators as shown in Table 5. This means that, given the prevalence of malnutrition and the logarithmic mean of consumption, the average severity of malnutrition among malnourished children within the same province tends to be higher in communes with higher inequality. This is consistent with Figure 6, because the average severity among malnourished children tends to increase with increased level of inequality regardless of the location of the threshold X on the horizontal axis.

If the effect of consumption on the prevalence of malnutrition is linear, inequality should have no effect on  $P^0$  of malnutrition. Hence, non-linear effects of consumption on nutrition are extremely important in understanding the relationship between consumption and malnutrition. This is particularly the case when we have only aggregate data. Prevalence is an important indicator, but we also need to look at the FGT measures with higher parameter values. DHS estimates are typically too aggregated to meaningfully investigate aggregate relationships. The small-area estimates of consumption and health are very helpful for exploring the relationship between poverty, inequality and malnutrition.

## 6.2 Decomposition analysis of health inequality

While the relationship between consumption poverty, inequality and the prevalence of malnutrition is important, inequality of child health, as measured by nutrition outcomes, is also

a matter of interest on its own. Health is an important aspect of human welfare, and health inequality is relevant to the equity of health, a broader notion (Sen, 2002). Given that child health indicators are important in predicting the future social standing of the child (Nystrom Peck, 1992; Montgomery et al., 1996), the existence of child health inequality in itself is indeed a concern.

In what follows, we shall focus on the decomposition analysis of inequality indicators. The analysis we carry out is different from most other studies on health inequality decomposition in at least three respects. First, despite the growing literature on health inequality, there have been relatively few studies carried out at a spatially disaggregated level, especially in developing countries. The methodology we developed in this study allows us to analyze health inequality at a level of small geographic areas even in the absence of a large number of observations of health indicators collected at that level.

Second, while the majority of health inequality literature uses mortality data, we use the standardized height and weight. Because death is a rare event, it may not be the best indicator at a very disaggregated level. Also, the probability of death is influenced by the individual's experiences over the course of his/her life. This makes it difficult to compare populations with different demographic groups. Standardized height and weight for children have some advantages here. Moreover, anthropometric indicators are likely to be more reliable than mortality data in countries like Cambodia where vital records are not fully maintained. Pradhan et al. (2003) discusses additional advantages of the standardized height measure.<sup>13</sup>

Third, we emphasize direct policy implications of inequality for targeting. Geographic targeting is meaningful only when there exists spatial heterogeneity in malnutrition or consumption. Hence, the between-location component of overall inequality is relevant to potential gains from geographic targeting. In many of the health inequality studies, the focus

---

<sup>13</sup>Pradhan et al. (2003) prefer height to weight because too much weight is obviously not good for health. Given that less than 1 percent of the children under five is overweight in Cambodia, this is a minor concern in this study.

is placed on the different health outcomes for different groups defined by the socioeconomic status (SES) or income groups (*e.g.* Kunst et al. (1998); van Doorslaer et al. (1997)). Such analyses provides us with a valuable description of the difference in health outcome among various groups in society. However, they are of little use for the purpose of targeting health programs, unless we can target resources on the basis of the SES or income. Even if such targeting is possible, it is still useful to combine it with geographic targeting.

Let us now proceed to the decomposition of health inequality. We present the results based on three decomposable indices. We begin with the approach proposed by Pradhan et al. (2003). We then propose two other decomposable measures of health inequality. One of them is the variance of FGT measures, and the other uses concentration curves for decomposition of the Gini index. Pradhan et al. (2003) apply  $GE(0)$  to the standardized height to decompose world health inequality into the between-country component and the within-country component. They adjust the proportion of between-group component by replacing the denominator (*i.e.*  $GE_{\text{World}}(0)$ ) with inequality in the reference population due to natural variation in height subtracted from the world health inequality (*i.e.*  $GE_{\text{World}}(0) - GE_{\text{Natural}}(0)$ ). They find that around 30 percent of total inequality is due to the between-country component. In a similar way, Table 7 provides the proportion of between-group inequality to total inequality adjusted for the natural inequality. Compared with global health inequality, the share of the between-group inequality is small in Cambodia. Table 7 also compares the share of between-group inequality for anthropometric indicators and consumption. Consumption is much more unevenly distributed across space than anthropometric indicators.<sup>14</sup> For the parameter values of the generalized entropy measure between -1 and 1, the share of the between-group com-

---

<sup>14</sup>Even though this point still holds, one should note that consumption does not capture intra-household inequality because the unit record of consumption data is a household. We can subtract within-household  $GE(0)$  from the denominator when calculating the share of between-group inequality in order to adjust for intra-household inequality. The numbers except for the last column in Table 7 will increase by only around 10 percent for both anthropometric indicators. This calculation ignores between-household inequality due to natural variation, and overestimates the share of between-group inequality. Accurate evaluation was not possible because we do not know the share of between-household and within-household components of natural inequality



ponent does not change much. For both anthropometric and consumption measures, the proportion of between-group inequality within Cambodia is much lower the proportion of between-country inequality in the world.<sup>15</sup>

The magnitude of within-group inequality varies quite substantially across the communes. Figure 7 shows the 95 percent confidence interval of  $GE(0)$  ranked from the most equal communes. As pointed out by Elbers et al. (2003b), a low share of between-group inequality is consistent with high heterogeneity in within-group inequality. This means that the levels of malnourishment across the communes may also be heterogeneous, even if the between-group inequality is small or even zero. Therefore, a low share of between-group  $GE(0)$  does not mean that geographic targeting is ineffective.

This point may be more clearly understood with a simple numerical example for income distribution. Suppose that there are two villages A and B in a small country and each village has 100 people. In Village A, everyone earns 10. In Village B, there is one rich person whose income is 901 and the remaining 99 people earn only 1. Let's suppose that the poverty line is 5. In this case, the between-group inequality measured by the generalized entropy measures is zero because the average income is 10 for both villages. However, geographic targeting is obviously very useful because all the poor people live in Village B. This shows that what really matters for the targeting of poverty alleviation programs is not the inequality of income but the inequality of poverty. Similarly, what matters for the targeting of child nutrition programs is the inequality of malnutrition, and not the inequality of height or weight.

To focus on the malnourishment, we propose another way of decomposing health inequality into between-group and within-group components. Let  $P$  be the FGT measure with an arbitrary parameter, and  $\mathcal{G}$  be a set of geographic groups such as communes, and  $\{\mathcal{I}_g\}_{g \in \mathcal{G}}$  forms a partition on the set of all the individuals  $\mathcal{I}$  in the population. We can decompose the variance  $V^T$  of  $P$  into within-group component  $V^W$  and between-group component  $V^B$

---

<sup>15</sup>Lovell (1998) decomposed  $GE(0)$  for consumption. About 70 percent of consumption inequality is explained by between-country inequality.

as follows:

$$V^T = \sum_{g \in \mathcal{G}} \sum_{i \in \mathcal{I}_g} (P_{gi} - \bar{P})^2 = \sum_{g \in \mathcal{G}} \sum_{i \in \mathcal{I}_g} (P_{gi} - \bar{P}_g)^2 + \sum_{g \in \mathcal{G}} N_g (\bar{P}_g - \bar{P})^2 = V^W + V^B$$

where  $N \equiv \#(\mathcal{I})$ ,  $N_g \equiv \#(\mathcal{I}_g)$ ,  $\bar{P} \equiv \frac{1}{N} \sum_{\mathcal{I}_g \in \mathcal{G}} \sum_{i \in g} P_{gi}$ , and  $\bar{P}_g \equiv \frac{1}{N_g} \sum_{i \in g} P_{gi}$

In this formulation, the proportion of the between-group variance to the total variance  $\frac{V^B}{V^T}$  is constant with respect to an affine transformation of the standardized height or weight. Hence, the arbitrary choice of reference sex and age group does not matter. Once the “poverty line” is accepted, the comparison of decomposition analysis for different indicators is easier with  $\frac{V^B}{V^T}$ . We looked at the between-group share of the variance of FGT measures with parameter 0, 1 and 2. Table 8 shows the results. As with the generalized entropy measures, the between group health inequality is lower for standardized height and weight than consumption.

One should note that  $\frac{V^B}{V^T}$  for FGT measures does not look at the entire distribution. Whether this is desirable property or not would depend on the purpose of the index. For the formulation of targeting policies we are more interested in the lower tail of the distribution. For example, a very tall child getting taller by hypothetically “stealing height” from a little less taller child would be much less concerning than a very short child getting shorter by having her height stolen from a less short child.

Another approach we propose to decompose inequality is by using the concentration curve and concentration index (CI) for an FGT measure  $P$ . This approach is attractive as it is directly related to targeting, a point discussed in more detail later. The concentration curve is a generalized version of the Lorenz curve, and is drawn according the group an individual belongs to rather than an individual ranking. Formally, the concentration curve is defined as follows. Let  $\mathcal{G} \equiv \{1, 2, \dots, G\}$ ,  $a_g \equiv \sum_{j=1}^g \frac{N_j}{N}$ , and  $b_g \equiv \sum_{j=1}^g \frac{\bar{P}_j N_j}{PN}$ . That is,  $a_g$  is the share

of the cumulative population and  $b_g$  the share of cumulative FGT measure for groups with index no greater than  $g$ . By definition, we have  $a_G = b_G = 1$ . The concentration curve  $C(q)$  is a piecewise linear function of the share of population  $q$ , which connects  $(0, 0)$  and  $(a_j, b_j)$  for  $\forall j \in \{1, 2, \dots, G\}$ . In general,  $\mathcal{G}$  may be defined by any individual characteristics, including SES and income groups. When  $P_g$  is not monotonic in  $g$ , the concentration curve may cross the 45-degree line. In this study, we focus on the case where  $\mathcal{G}$  represents geographic groups. In addition, we assume that the geographic groups are sorted by the descending order of  $P$  so that  $P_j \leq P_k$  for  $\forall j, k \in \mathcal{G}$  such that  $k \leq l$ . In this case, the concentration curve is concave and does not cross the 45-degree line. When  $\mathcal{G}$  represents communes, we shall call  $C(q)$  the commune-level concentration curve, and use a similar terminology for other levels of geographic aggregation. The individual-level concentration curve ( $C(q)$  for  $\mathcal{G} = \mathcal{I}$ ) is the Lorenz curve for  $P$  rotated by 180 degrees.

Figure 8 gives an example of a concentration curve for  $P^0$ . The horizontal axis measures the cumulative population share  $q$  and the vertical axis the cumulative share  $p$  of  $P^0$  with  $\overline{OA} = \overline{OC} = 1$ . At the individual level,  $P^0$  can take only zero or one. Hence the individual-level concentration curve looks like the bold line OJB. Since  $\overline{JB}$  is the proportion of the well-nourished children to all the children,  $\overline{CJ}$  is the prevalence of malnutrition  $P^0$  for the entire population. The Gini index in Figure 8 is twice the area of the triangle OJB, which is the area above the 45-degree line and under the individual-level concentration curve. When the parameter for the FGT measure is zero, the Gini index is  $1 - P^0$ . The line OJ is straight for  $P^0$ , but may not be straight when  $P$  has a parameter greater than 0.

Now, consider the commune-level concentration curve. If we have three communes with different levels of malnutrition, the concentration curve would look like the line OFIB. The concentration index is twice the area under the 45-degree line and above the concentration curve, that is  $CI(= 1 - 2 \int_0^1 C(q) dq)$ . This takes a negative number when the concentration curve lies above the 45-degree line. The absolute value of CI is twice the lightly shaded area

OFIB in Figure 8. The Gini index is the CI when the concentration curve is the Lorenz curve, and provides the upper bound for the maximum absolute value of the CI. Several decomposition techniques of CI have been proposed. (Clarke et al., 2003; Wagstaff et al., 2003; Wagstaff and van Doorslaer, 2004) We propose to decompose the Gini index for FGT measures using an approach very similar to the one used in Wagstaff and van Doorslaer (2004), First we decompose the Gini index  $GI$  into between-group negative concentration index  $NCI_B$ , within-group negative concentration index  $NCI_W$  and the residual  $GI_R$  so that  $GI = NCI_B + NCI_W + GI_R$ . This decomposition corresponds to the change in negative CI for a hypothetical re-ordering and redistribution in four steps. First, suppose everyone is sorted by the descending order of the FGT measure so that the concentration curve  $C_1$  is OJB. Second, suppose that we sort the individuals by the descending order of the commune-level mean FGT measure each individual belongs to, while keeping the relative ranking within the commune. The resulting concentration curve  $C_2$  looks like OEFHIKB. Third, we redistribute FGT measure so that everyone gets the commune-level average, in which case the concentration curve  $C_3$  is OFIB. Finally, we redistribute across the communes so that everyone gets the population-average FGT and the concentration curve  $C_4$  becomes the 45-degree line OB.

$NCI_B$  is twice the area between  $C_3$  and  $C_4$ , or the lightly shaded area OFIB in Figure 8.  $NCI_W$  is twice the area between  $C_2$  and  $C_3$ , or the sum of three triangles OEF, FHI and IKB, which are heavily shaded in Figure 8. Finally, the residual  $GI_R$  is twice the area between  $C_1$  and  $C_2$ , or EFHIKJ.  $GI_R$  is analogous to the transvariation component of the Gini decomposition. Table 9 shows the results of decomposition of the Gini index for standardized height. As with GE(0) and the variance of FGT, the proportion of the between-group component is not large in absolute value. As with other decomposable indicators, the proportion of the between-group inequality  $\frac{NCI_B}{GI}$  was similar for standardized weight, and higher for consumption.

The decomposition analysis based on Gini and CI is not as neat as the decomposition of  $GE(0)$  or  $V^T$  because of the existence of  $GI_R$ . The definition of within-group is somewhat *ad hoc* because splitting one commune into two communes with the same distribution of the FGT measure will reduce the within-group component and increase the residual. Yet the CI is suitable for examining the magnitude of the between-group inequality and has some unique attractions because of its direct relation to the potential gains from geographic targeting. While we have seen that the share of between-group inequality is not very large, it has not been clear whether geographic targeting can be still effective. We now try to answer this question.

### 6.3 Concentration curve and geographic targeting

Let us now interpret the concentration curve in the context of geographic targeting. Suppose that we want to reduce the total FGT measure by the proportion  $\overline{OG}$  in Figure 8. Because everyone is malnourished on the portion OJ in the individual-level concentration curve, we need to bring the proportion  $\overline{OD}$  of the total population out of malnourishment. Hence, the vertical axis can be interpreted as the rate of reduction in FGT, and the horizontal axis can be interpreted as the proportion of the population targeted.

In this setup, we assume that those who receive assistance are completely brought out of malnourishment. If the *per capita* cost of assistance is fixed and normalized to the reciprocal of the total population, the horizontal axis also represents the cost of FGT reduction. Fixed *per capita* cost may be a reasonable approximation when the child nutrition program involves direct aid in fixed amount in a relatively large scale. Distribution of micronutrients may fall in this category. We shall relax this assumption in the subsequent discussion.

Suppose now that the FGT parameter is greater than zero so that the concentration curve looks like in Figure 9(a). In this figure,  $\overline{OE}$  is the proportion of total malnourished children to the total population (*i.e.*  $P^0$ ). The dashed line OKB is the commune-level concentration

curve.

Let us suppose that the policy-maker wants to eliminate malnutrition by a child nutrition program. However, the policy maker can reduce the FGT measure only by the proportion  $\overline{OI}$  because the resources available for the program are limited. Then, noting that  $\overline{OD}$  is the proportion of malnourished children included in the program to the total population, the rate of Type I error, or the error of exclusion of malnourished children in the program, is  $\frac{\overline{DE}}{\overline{OE}}$ . When the parameter for the FGT measure is zero, it coincides with  $\overline{IC}$ . If perfect individual-level targeting is possible, the cost of reducing the FGT measure by the proportion  $\overline{OI}$  is  $\overline{OD}$ , because it is the proportion of the beneficiaries of the program to the total population.

Now suppose instead that we can target the program at the commune level. By this, we mean we can distinguish children if and only if they are in different communes. We also assume that the program does not induce children to move, which is a reasonable assumption because the cost of changing the place of residency is very high for poor people. In general, not everyone in the least malnourished commune included in the targeting policy receives the program. We assume that the program in the least malnourished commune is targeted uniformly, that is, everyone in that commune receives the program with the same probability.<sup>16</sup> With the commune-level targeting, the cost of achieving the same goal in expectation is  $\overline{OF}$  and the leakage of the program to well-nourished children is  $\overline{DF}(= \overline{OF} - \overline{OD})$ . Therefore, the rate of the Type II error, or the error of inclusion of well-nourished children in the program, is  $\frac{\overline{DF}}{\overline{OF}}$  in this case.

While both Type I and Type II errors are closely related to the efficiency of targeting, they are not very satisfactory measures of efficiency. Type I error tends to go down as the goal of FGT reduction or the budget available for the program is high, regardless of the level of geographic areas at which targeting is conducted. Type II error is also slightly difficult

---

<sup>16</sup>Recall that *per capita* cost of the program is fixed. Suppose an assistance worth one dollar goes to a commune of hundred individuals in the least malnourished commune included in the targeting policy. Instead of each individual receiving one cent for sure, each individual receives one dollar with probability 0.01.

to interpret because a comparison is made against the perfect targeting. We do not exactly know from the Type II error how much improvement can be made by introducing a particular targeting scheme.

We use two types of efficiency measures which we shall call *budgetary gains* and *equivalence gains*. The budgetary gains measure how much total expenditure can be reduced to achieve a given level of malnutrition reduction in comparison with the reference case. The equivalence gains measure how much more malnutrition reduction one can achieve with a fixed budget compared with the reference case. As with of Elbers et al. (2004b), we take uniform targeting at the national level as the reference case. That is, everyone receives the program with an equal probability in the reference case, and the corresponding national-level concentration curve is the 45-degree line OB in Figure 9(a).<sup>17</sup>

Because the cost of achieving the reduction  $\overline{OI}$  in Figure 9(a) by perfect (individual-level), commune-level and uniform (national-level) targeting are  $\overline{OD}$ ,  $\overline{OF}$  and  $\overline{OG}$ , the equivalence gains are respectively  $\overline{DG}$ ,  $\overline{FG}$  and 0. Similarly, given the budget  $\overline{OF}$ , the malnutrition reduction one can achieve from individual-level, commune-level and uniform targeting are  $\overline{OC}(= 1)$ ,  $\overline{OI}$  and  $\overline{OH}(= \overline{OF})$  respectively. Hence, the equivalence gains from individual-level, commune-level and uniform targeting are  $\overline{HC}$ ,  $\overline{HI}$  and 0 respectively. Both budgetary gains and equivalence gains take a value between 0 and 1 for the concentration curves we consider. They can take negative values if the targeting policy favors well-nourished children over malnourished children.

Budgetary gains and equivalence gains measure appropriately the performance of a particular targeting policy. However, we would be also interested in the potential gains from targeting are captured by a particular targeting policy. In Figure 9(a), commune-level targeting leaves considerable room for improvement and thus the policy maker may want to

---

<sup>17</sup>The concentration curve we calculate from the commune-level estimates are subject to the statistical errors. Since the calculation of FGT reduction is made at the national level, the standard errors are usually very small.

combine geographic targeting with other forms of targeting such as self-targeting.<sup>18</sup> However, if the individual-level concentration curve were OJ'B instead of OJB, there would be almost no room for improvement. In this case, the policy-maker will not gain much from devising a more complex targeting scheme.

We propose *relative budgetary gains* to measure how much a particular targeting policy is capturing the possible gains from targeting. In Figure 9(a), the relative budgetary gain from community-level targeting given the goal of malnutrition reduction  $\overline{OI}$  is  $\frac{\overline{FG}}{\overline{DG}}$ . By definition, the *relative budgetary gain* is 1 for perfect targeting and 0 for uniform targeting regardless of the goal of FGT reduction.

There may not be an obvious choice of the goal  $\overline{OI}$  or the budget  $\overline{OF}$ . It is convenient to have an overall measure of the efficiency measure of targeting. A natural choice would be the negative concentration index. For the commune-level targeting, it is twice the heavily shaded area X in Figure 9(a). Notice that  $2 \cdot X = \frac{X}{\overline{X} + \overline{Y} + \overline{Z}}$ . This measure is the budgetary gain averaged over all the possible values of FGT reduction goals. Hence, we shall call this *average budgetary gains*. An alternative interpretation of the average budgetary gains is the equivalence gains averaged over the interval between zero and one. Similarly, we shall call the ratio of negative concentration index for commune-level targeting to that for perfect targeting *average relative budgetary gains*, which is  $\frac{X}{\overline{X} + \overline{Y}}$  in Figure 9(a). It is also equal to  $\frac{NCI_B}{GI}$ , or the between-group negative concentration index over the Gini index.

Using the point estimates of the prevalence of stunting and underweight, we drew the concentration curves for different levels of geographic aggregation as shown in Figure 10(a). The concentration curves for underweight look similar and are given in Figure 10(b). There are four points to note here. First, the commune-level targeting is a large improvement over the stratum-level targeting, which is the best policy-makers can do with the DHS data only.

---

<sup>18</sup>An example of a self-targeting program is the food-for-work program, because only poor people without better outside option are willing to participate in the program to receive food. Food-for-work targeted to a particular region is an example of the combination of geographic targeting and self-targeting.



As shown in Table 9, the *average relative budgetary gains*  $\frac{NCI_B}{GI}$  at the commune level is more than two times larger than that at the provincial level. When compared with ecozone-level or stratum-level targeting, the improvement is even more dramatic.<sup>19</sup>

Second, on average, there remains large room for improvement even if targeting is carried out at the level of very small geographic areas such as the village level. Hence, it is very important for policy-makers to combine geographic targeting with other forms of targeting whenever feasible. Third, the equivalence gains vary substantially depending on the proportion of targeted people. This point can be seen more clearly in Figure 11(a), which shows the equivalence gains for different levels of geographic targeting. We see that the equivalence gain first goes up and then goes down. By definition, equivalence gains take zero when the proportion of targeted population is zero or one. The equivalence gains for underweight is given in 11(b).

Fourth, as with the equivalence gains, budgetary gains change with the goal of FGT reduction. Relative budgetary gains are suitable for evaluating how well particular targeting scheme is for a given goal of FGT reduction. Figure 12(a) shows the relative budgetary gains for stunting. Clearly, commune-level targeting is more efficient if the goal of FGT reduction is lower, capturing most of the potential gains from targeting. When the goal is low, commune-level targeting can be as high as five times more efficient than stratum-level targeting. This is an important point, because realistic values of FGT reduction in the short run will not be extremely high. For example, a Cambodia Millennium Development Goal is to reduce the prevalence of stunted children by 35 percent by 2010 and 50 percent by 2015 (MoP, 2003). Hence, the *average* relative budgetary gains somewhat understate the efficiency gains of commune-level targeting in the short run. The relative budgetary gains

---

<sup>19</sup>To make the graphs readable, we omit stratum-level targeting in the graphs. Stratum-level targeting is close to provincial-level targeting but slightly less efficient because it is done at a more aggregated level. We prefer to present provincial-level estimates because they are practically more important. Also, while we report the results for village-level targeting, we take the commune-level targeting as a benchmark for disaggregated geographic targeting. This is because commune-level targeting is more realistic than village-level targeting as communes have some degrees of autonomy whereas villages do very little.

for underweight are given in Figure 12(b).

As we have already noted, the efficiency measures we used are based on the assumption that the *per capita* cost of the program is fixed. Hence, the concentration curve is useful when the cost of targeting is approximately in proportion to the share of the population covered by targeting. However, if the cost of improving the nutrition status differs across the individuals depending on how severely the individual is malnourished, we need to take into account the variable costs. This is particularly important when we use higher values of the parameter in the FGT measures.

To this end, we consider the other extreme case where the *per capita* cost is variable. We assume that increasing one unit of a standardized anthropometric indicator of a child costs the same for all the children. This assumption would be reasonable for direct food (macronutrient) aid programs. We also assume in the subsequent discussion that the parameter  $\theta$  of the FGT measure of interest satisfies  $\theta \geq 1$ . Otherwise, it is optimal to target assistance to children whose height or weight is just under the level of negative two Z-score, but such a targeting policy is unethical and hard to justify.

Now, we shall propose the *adjusted* concentration curve, which takes into account the different magnitude of malnourishment in the population. Instead of the share of the population, we have the budget in the horizontal axis, as measured by the total increase in the standardized height or weight. We shall normalize it by the expected budget that is required to eliminate malnutrition, which is equal to the point estimate of total malnutrition gap  $P^1 \cdot N$ . Given the budget and the level of aggregation at which targeting is carried out, we calculate the optimal level of transfers to each location and find the expected reduction in the FGT measure. Plotting the share of the FGT measure reduced in expectation against various levels of budget, we obtain the *adjusted* concentration curve. Figure 9(b) provides an illustration of adjusted concentration curves. It is similar to Figure 9(a) in many respects, and we can indeed define *budgetary gains* and *equivalence gains* in exactly the same manner.

However, there are four differences to note. First, the curve is not piecewise linear because the marginal impact of the child nutrition program diminishes even within the same geographic location as more and more children get less malnourished. Second, because of the first point, the adjusted concentration curve for the reference case of uniform targeting is not the 45-degree line any more. In Figure 9(b), the concentration curve for uniform targeting is OLB. Third, the scale of the horizontal axis is different. In Figure 9(a),  $\overline{OA} = 1$ , but in Figure 9(b), we do not have such a simple relationship.

Fourth, the standard errors associated with the point estimates play more important role for the adjusted concentration curve. For example, when the individual-level concentration curve is based on data and not estimates, we have  $\overline{OE} = 1$  in Figure 9(b). If it is based on estimates,  $\overline{OE} \geq 1$ . This is because eliminating malnutrition with probability one costs more than the expected cost of elimination of malnutrition.

Also, in the previous setting where the *per capita* cost was fixed, it was optimal to target from the locations with highest point estimate of  $P^0$ . Since the assumption was such that the proportion of  $P^0$  reduced in the location is equal to the share of the population targeted in that location, we only needed to use the point estimate. However, the adjusted concentration curve requires more complex calculation because we need to know how the FGT measures change according to different levels of transfers.

To draw adjusted concentration curves, we need to find the value of minimum expected FGT measure given the budget and the level of geographic targeting. We use the imputed anthropometric indicators  $\tilde{y}$  from previous sections. Formally, given the budget  $B$ , the minimization problem the policy-maker solves is:

$$\begin{aligned}
& \min_{\{t_g\}_{g \in \mathcal{G}}} \frac{1}{NR} \sum_{r=1}^R \sum_{g \in \mathcal{G}} \sum_{i \in \mathcal{I}_g} \left( \frac{z - \tilde{y}_{gi,(r)} - t_g}{z} \right)^\theta \cdot \text{Ind}(z > \tilde{y}_{gi,(r)} + t_g) \\
& \text{s.t. } \sum_{g \in \mathcal{G}} t_g \leq B \quad \text{and} \quad \forall g \in \mathcal{G}, t_g \geq 0
\end{aligned}$$

where  $t_g$  is the amount of transfer of the standardized height or weight expressed as a proportion of the total malnutrition gap. Let the set of targeted locations be  $\mathcal{T}(\equiv \{g \in \mathcal{G} | t_g > 0\})$ . It is straightforward to establish that the following condition is necessary and sufficient for optimality when  $\theta > 1$ . For  $\forall \tau \in \mathcal{T}$  and  $\forall h \in \mathcal{G} - \mathcal{T}$ , we have  $\exists \bar{P}^{\theta-1}$  such that  $P_\tau^{\theta-1} = \bar{P}^{\theta-1}$  and  $P_h^{\theta-1} \leq \bar{P}^{\theta-1}$ . This means that there is some value  $\bar{P}^{\theta-1}$  of the FGT measure with parameter  $\theta - 1$ , which is common among the targeted locations and no smaller than that in any other locations.

This is a useful result for calculating the adjusted concentration curve. We employ a computational algorithm very similar to the one used in Elbers et al. (2004b). First, we can pick various values of  $\bar{P}^{\theta-1}$ . We then calculate  $P_g^{\theta-1}$  for various levels of transfer for each  $g \in \mathcal{G}$ , and linearly interpolate to find the transfer amount  $t_g$  to achieve the given level of  $\bar{P}^{\theta-1}$ . If the pre-transfer  $P^{\theta-1}$  is less than  $\bar{P}^{\theta-1}$ , we have  $t_g = 0$ . Once we find the transfer for each location, we can calculate the budget and post-transfer  $P^\theta$ .

Figure 13(a) and Figure 13(b) show the adjusted concentration curves for stunting and underweight for  $\theta = 2$ . The pattern of the adjusted concentration curve is qualitatively somewhat similar to Figure 10(a) and Figure 10(b) in that the individual-level concentration curve is in the far left of other curves, suggesting that there remains large room for improvement even if targeting is conducted at the commune level or village level. It also suggests that commune-level targeting is a significant improvement from provincial-level or stratum-level targeting as is clear from the graph of relative budgetary gains shown in Figure 14. The ratio of average relative budgetary gains for commune-level and provincial-level targeting exceeds

four for both stunting and underweight.

We have considered two extreme cases, one where the cost *per capita* is fixed, and the other where it varies with the individual level of malnutrition. In both cases, we found that geographic targeting alone leaves large room for improvement. Clearly, geographic targeting is not a panacea for maximizing the impact of limited resources. However, this does not undermine the usefulness of geographic targeting because geographic targeting can be easily combined with other forms of targeting in most cases in order to further improve the efficiency. In fact, we found that commune-level targeting alone improves efficiency by more than two times compared with the stratum-level targeting. The improvement is much more dramatic when the goal is a modest reduction. Given this, the commune-level estimates we produced can indeed substantially improve the efficiency of targeting.

## 7 Conclusion

Estimates of the prevalence of child malnutrition were previously available only at the level of CDHS strata. Stratum-level estimates often mask great disparities in the prevalence of malnutrition within the stratum. Unless there are strata with an extremely high prevalence of malnutrition, targeting based on stratum level estimates is unlikely to capture many of the malnourished children and likely to misallocate resources.

To overcome the problem of limited data, we have developed a methodology to estimate the prevalence of child malnutrition at the level of small geographic areas. We disaggregated the estimates of the prevalence of child malnutrition in Cambodia from currently available 17 CDHS strata into 1,594 communes. We have extended the ELL small-area estimation technique to jointly estimate multiple indicators and allow for a richer structure of error terms. This is a crucial step to address issues unique to nutrition indicators. While we applied this methodology to the Cambodian data, it can be easily applied to other countries

where census data and survey data with an anthropometric component are available.

Although estimated standard errors in our study are quite high for some communes, the magnitude of the standard errors for the estimated prevalence of stunting and underweight at the commune level is, on average, comparable to that for the existing stratum-level estimates derived only from the CDHS data. We argued that high levels of standard errors for each commune are not necessarily worrisome when a proposed policy intervention delivers assistance to a relatively large number of communes.

The commune-level estimates we derived in this study can be used for analyzing the current situation of malnutrition and for shaping targeting policies. We have presented the commune-level estimates as nutrition maps. These maps allow policy-makers to visually locate areas of severe malnutrition. We argued that overlaid maps help us identify possible causes of malnutrition in different locations. This in turn provides policy-makers with valuable information on the appropriate design of child nutrition programs. For example, direct food aid would be helpful if food shortage is the principal cause of malnutrition. However, local health clinics may prove more helpful where many children suffer from diseases such as malaria and diarrhea.

We demonstrated three applications of the commune-level estimates. First, we applied them to the analysis of correlation between consumption poverty, inequality and malnutrition. We found no simple relationship, but the regression results suggest that non-linear effects of consumption on the child nutrition status are crucial for understanding the relationship. Our small-area estimates of consumption and health are very helpful for exploring the relationship at a disaggregated level between malnutrition and other indicators, including consumption poverty and inequality.

Second, we conducted decomposition analyses of child health inequality. We first followed the approach by Pradhan et al. (2003), and then proposed two additional ways of decomposing inequality by groups. One of them is based on the analysis of variance, and the other uses

concentration curves. We consistently found that the between-group component is smaller for health inequality than consumption inequality. We also found that the proportion of between-group inequality in the country is small in comparison with the proportion of between-country inequality in the world. While this application is purely descriptive, it is useful for elucidating the significance of the geographic information in explaining overall inequality. The magnitude of the ratio of between-location health inequality to the overall health inequality was not previously known.

Finally, we evaluated the efficiency gains from geographic targeting. We showed that concentration curves are closely linked to efficiency measures of targeting. We proposed the *adjusted* concentration curve to allow for the variable transfer across the individuals. We found that commune-level targeting based on the estimates we produced is on average at least two to three times more efficient than stratum-level targeting. This is the best we can do with the CDHS estimates. The budgetary gains for commune-level targeting can be five times as high as those for stratum-level targeting in the short run.

All three of these applications illustrate interesting research questions which can now be answered with the commune-level estimates we derived. Successful application of our methodology in other countries is likely to help improve the efficiency of targeting in these countries. It will also help us to understand the relationship between consumption inequality and health, and inequality in health outcomes.

## References

- Alderman, H. (2000) ‘Anthropometry.’ In *Designing Household Survey Questionnaires for Developing Countries: Lessons from Ten Years of LSMS Experience*, ed. M. Grosh and P. Glewwe (Oxford University Press) pp. 251–272
- Alderman, H., M. Babita, G. Demombynes, N. Makhatha, and B. Özler (2002) ‘How small can you go? combining census and survey data for mapping poverty in South Africa.’ *Journal of African Economies* 11, 169–200

- Clarke, P.M., U. Gerdtham, and L.B. Connelly (2003) 'A note on the decomposition of the health concentration index.' *Health Economics* 12, 511–516
- Curtis, S.L., and M. Hossain (1998) 'West Africa spatial analysis prototype explanatory analysis: The effects of aridity zone on child nutritional status.' Technical Report, Demographic and Health Survey, Macro International
- de Onis, M., E.A. Frongillo, and M. Blössner (2000) 'Is malnutrition declining? an analysis of changes in levels of child malnutrition since 1980.' *Bulletin of the World Health Organization* 78, 1222–1233
- Deaton, A. (2003) 'Health, inequality, and economic development.' *Journal of Economic Literature* XLI, 113–156
- Demombynes, G., J. Elbers, C. and Lanjouw, J. Lanjouw, J. Mistiaen, and B. Özler (2002) 'Producing an improved geographic profile of poverty: Methodology and evidence from three developing countries.' WIDER Discussion Paper 2002/39, United Nations University
- Dibley, M.J., J.B. Goldsby, N.W. Staehling, and F.L. Trowbridge (1987a) 'Development of normalized curves for the international growth reference: historical and technical considerations.' *American Journal of Clinical Nutrition* 46(5), 736–48
- Dibley, M.J., N.W. Staehling, P. Nieburg, and F.L. Trowbridge (1987b) 'Interpretation of Z-score anthropometric indicators derived from the international growth reference.' *American Journal of Clinical Nutrition* 46(5), 749–762
- Elbers, C., J.O. Lanjouw, and P. Lanjouw (2000) 'Welfare in villages and towns: Micro-level estimation of poverty and inequality.' Timbergen Institute Discussion Paper TI 2000-029/2, Timbergen Institute
- (2002) 'Micro-level estimation of welfare.' Policy Research Department Working Paper 2911, The World Bank
- (2003a) 'Micro-level estimation of poverty and inequality.' *Econometrica* 71(1), 355–364
- (2004a) 'Imputed welfare estimates in regression analysis.' *Journal of Economic Geography*
- Elbers, C., P. Lanjouw, J. Mistiaen, B. Özler, and K. Simler (2003b) 'Are neighbours equal? estimating local inequality in three developing countries.' FCND Discussion Paper No. 147, Food Consumption and Nutrition Division, International Food Policy Research Institute
- Elbers, C., T. Fujii, P. Lanjouw, B. Özler, and W. Yin (2004b) 'Poverty alleviation through geographic targeting.' Policy Research Working Paper 3419, The World Bank
- Foster, J.E., J. Greer, and E. Thorbecke (1984) 'A class of decomposable poverty indices.' *Econometrica* 52, 761–766



- Frongillo, E.A., M. de Onis, and K.M.P. Hanson (1997) 'Socioeconomic and demographic factors are associated with worldwide patterns of stunting and wasting of children.' *Journal of Nutrition* 127(12), 2302–2309
- Fujii, T. (2004) 'Commune-level estimation of poverty measure and its application in Cambodia.' In *Spatial Disparities in Human Development: Perspectives from Asia*, ed. Kanbur R., Venables A.J., and G. Wan (United Nations University Press). Forthcoming
- Galler, J.R., and L.R. Barrett (2001) 'Children and famine: Long-term impact on development.' *Ambulatory Child Health* 7, 85–95
- Glewwe, P., H.G. Jacoby, and E.M. King (2001) 'Early childhood nutrition and academic achievement: a longitudinal analysis.' *Journal of Public Economics* 81, 345–368
- Haddad, L., and R. Kanbur (1990) 'How serious is the neglect of intra-household inequality.' *The Economic Journal* 100(402), 866–81
- Hardy, F., and Health Unlimited Ratanakiri Team (2001) 'Health situation analysis Ratanakiri, Cambodia.' Technical Report, Health Unlimited
- Haughton, D., and J. Haughton (1997) 'Explaining child nutrition in Vietnam.' *Economic Development and Cultural Change* 45(3), 541–556
- Hentschel, J., J.O. Lanjouw, P. Lanjouw, and J. Poggi (2000) 'Combining census and survey data to study spatial dimensions of poverty: A case study of Ecuador.' *The World Bank Economic Review* 14(1), 147–166
- Kakwani, N. (1977) 'Applications of lorenz curves in economic analysis.' *Econometrica* 45(3), 719–728
- Kakwani, N., A. Wagstaff, and E. van Doorslaer (1997) 'Socioeconomic inequalities in health: Measurement, computation, and statistical inference.' *Journal of Econometrics* 77, 87–103
- Kanbur, R. (1987) 'Transfers, targeting and poverty.' *Economic Policy* 4(1), 112–147
- Kawachi, I., B.P. Kennedy, K. Lochner, and D. Prothrow-Stith (1997) 'Social capital, income inequality and mortality.' *American Journal of Public Health* 87(9), 1491–1498
- Khorshed Alam Mozumder, A.B.M., T.T. Kane, A. Levin, and S. Ahmed (2000) 'The effect of birth interval on malnutrition in bangladeshi infants and young children.' *Journal of Biosocial Science* 32, 289–300
- Kunst, A.E., F. Groenhof, J.P. Mackenbach, and EU Working Group on Socioeconomic Inequalities in Health (1998) 'Occupational class and cause specific mortality in middle aged men in 11 european countries: comparison of population based studies.' *British Medical Journal* 316, 1636–1642

- Li, Y., G. Guo, A. Shi, Y. Li, T. Anne, and H. Ushijima (1999) 'Prevalence and correlates of malnutrition among children in rural minority areas of China.' *Pediatrics International* 41, 549–556
- Lovell, M.C. (1998) 'Inequality within and among nations.' *Journal of Income Distribution* 8(1), 5–44
- Mahalanobis, P.C. (1960) 'A method of fractile graphical analysis.' *Econometrica* 28(2), 325–351
- Monteiro, C.A., L. Mondini, A.M. Torres, and I.M. dos Reis (1997) 'Patterns of intra-familial distribution of undernutrition: Methods and applications for developing societies.' *European Journal of Clinical Nutrition* 51, 800–803
- Montgomery, S.M., M.J. Bartley, D.G. Cook, and W.E.J. Wadsworth (1996) 'Health and social precursors of unemployment in young men.' *Journal of Epidemiology and Community Health* 50, 415–422
- MoP (2003) 'Cambodia millennium development goals report 2003.' Technical Report, Ministry of Planning, Royal Government of Cambodia
- National Institute of Statistics, Directorate General for Health, and ORC Macro (2001) *Cambodia Demographic and Health Survey 2000* (National Institute of Statistics, Directorate General for Health and ORC Macro)
- Nystrom Peck, A.M. (1992) 'Childhood environment, intergenerational mobility, and adult health—evidence from Swedish data.' *Journal of Epidemiology and Community Health* 46, 71–74
- Pelletier, D.L., E.A. Frongillo Jr., D.G. Schroeder, and J-P. Habicht (1994) 'A methodology for estimating the contribution of malnutrition to child mortality in developing countries.' *Journal of Nutrition* 124, 2106S–2122S
- Pradhan, M., D.E. Sahn, and S.D. Younger (2003) 'Decomposing world health inequality.' *Journal of Health Economics* 22(2), 271–293
- Ravallion, M. (1988) 'Expected poverty under risk-induced welfare variability.' *The Economic Journal* 98(393), 1171–1182
- Ravallion, M., and K. Chao (1989) 'Targeting policies for poverty alleviation under imperfect information: Algorithms and applications.' *Journal of Policy Modeling* 11(2), 213–224
- Sen, A. (2002) 'Why health equity?' *Health Economics* 11, 659–666
- Shariff, Z.M., J.T. Bond, and N.E. Johnson (2000) 'Nutrition and educational achievement of urban primary schoolchildren in Malaysia.' *Asia Pacific Journal of Clinical Nutrition* 4(9), 264–273

- Shorrocks, A.F. (1980) ‘The class of additively decomposable inequality measures.’ *Econometrica* 48(3), 613–626
- van Doorslaer, E., A. Wagstaff, H. Bleichrodt, S. Calonge, U. Gerdtham, M. Gerfin, J. Geurts, L. Gross, U. Häkkinen, R.E. Leu, O. O’Donnell, C. Propper, F. Puffer, M. Rodríguez, G. Sundberg, and O. Winkelhake (1997) ‘Income-related inequalities in health: some international comparisons.’ *Journal of Health Economics* 16(1), 93–112
- Victora, C. (1992) ‘The association between wasting and stunting: An international perspective.’ *Journal of Nutrition* 122(5), 1105–1110
- Wagstaff, A., and E. van Doorslaer (2000) ‘Income inequality and health: What does the literature tell us?’ *Annual Review of Public Health* 21, 543–567
- (2004) ‘Overall versus socioeconomic health inequality: a measurement framework and two empirical illustrations.’ *Health Economics* 13, 297–301
- Wagstaff, A., E. van Doorslaer, and N. Watanabe (2003) ‘On decomposing the causes of health sector inequalities with an application to malnutrition inequalities in Vietnam.’ *Journal of Econometrics* 112, 207–223
- Waterlow, J.C., R. Buzina, W. Keller, J.M. Lane, M.Z. Nichaman, and J.M. Tanner (1977) ‘The presentation and use of height and weight data for comparing the nutritional status of groups of children under the age of 10 years.’ *Bulletin of the World Health Organization* 55(4), 489–498
- WFP (2001) *Identifying Poor Areas in Cambodia: Combining Census and Socio-Economic Survey Data to Trace the Spatial Dimensions of Poverty* (Phnom Penh: World Food Programme)
- WHO Working Group (1986) ‘Use and interpretation of anthropometric indicators of nutritional status.’ *Bulletin of the World Health Organization* 64(6), 929–941
- (1995) ‘An evaluation of infant growth: the use and interpretation of anthropometry in infants.’ *Bulletin of the World Health Organization* 73(2), 165–174
- Wilkinson, R.G. (1996) *Unhealthy Societies: The Afflictions of Inequality* (Routledge)
- World Health Organization (2002) *World Health Report 2002: Reducing risks and promoting healthy life* (Geneva: World Health Organization)
- Yitzhaki, S., and J. Slemrod (1991) ‘Welfare dominance: An application to commodity taxation.’ *American Economic Review* 81(3), 480–496
- Zeini, L.O., and J.B. Casterline (2002) ‘Clustering of malnutrition among Egyptian children.’ mimeo, Cairo University and Population Council

## Appendix

### A Mathematical Proofs

**Proof of Eq (1)** First note the following:

$$\begin{aligned} E[(u_{ch.}^{(k)})^2] &= (\sigma_\eta^{(k)})^2 + (\sigma_{\epsilon, ch}^{(k)})^2 + \frac{1}{I_{ch}} (\sigma_\delta^{(k)})^2 \\ E[(u_{c..}^{(k)})^2] &= (\sigma_\eta^{(k)})^2 + \frac{1}{H_c^2} \sum_{h \in \mathcal{H}_c} (\sigma_{\epsilon, ch}^{(k)})^2 + \frac{1}{H_c^2} \left( \sum_{h \in \mathcal{H}_c} \frac{1}{I_{ch}} \right) (\sigma_\delta^{(k)})^2 \end{aligned}$$

Hence,

$$\begin{aligned} & \sum_{c \in \mathcal{C}} \frac{w_c}{H_c} \sum_{h \in \mathcal{H}_c} E[(u_{ch.})^2] \\ &= \sigma_\eta^2 + \sum_{c \in \mathcal{C}} w_c H_c \left( \frac{1}{H_c^2} \left( \sum_{h \in \mathcal{H}_c} \sigma_{\epsilon, ch}^2 \right) + \frac{1}{H_c^2} \left( \sum_{h \in \mathcal{H}_c} \frac{1}{I_{ch}} \right) \sigma_\delta^2 \right) \\ &= \sigma_\eta^2 + \sum_{c \in \mathcal{C}} w_c H_c (E[(u_{c..})^2] - \sigma_\eta^2) \\ &= (1 - \sum_{c \in \mathcal{C}} w_c H_c) \sigma_\eta^2 + \sum_{c \in \mathcal{C}} w_c H_c E[(u_{c..})^2] \end{aligned}$$

Solving for  $\sigma_\eta^2$ , we have Eq(1).

**Proof of Eq (2)** First note the following:

$$E[(u_{ch.} - u_{c..})^2] = \frac{H_c - 2}{H_c} \left\{ \frac{\sigma_\delta^2}{I_{ch}} + \sigma_{\epsilon, ch}^2 \right\} + \frac{1}{H_c^2} \left\{ \sum_{h' \in \mathcal{H}_c} \left( \frac{\sigma_\delta^2}{I_{ch'}} + \sigma_{\epsilon, ch'}^2 \right) \right\}$$

By summing over the households in each cluster, we have

$$E\left[ \sum_{h' \in \mathcal{H}_c} (u_{ch'.} - u_{c..})^2 \right] = \frac{H_c - 1}{H_c} \left\{ \sum_{h' \in \mathcal{H}_c} \left( \frac{\sigma_\delta^2}{I_{ch'}} + \sigma_{\epsilon, ch'}^2 \right) \right\}$$

Therefore, for all the households in  $\mathcal{C}^*$ ,

$$\sigma_{\epsilon, ch}^2 = \frac{H_c}{H_c - 2} E[(u_{ch\cdot} - u_{c\cdot\cdot})^2] - \frac{1}{(H_c - 1)(H_c - 2)} E\left[\sum_{h' \in \mathcal{H}_c} (u_{ch'\cdot} - u_{c\cdot\cdot})^2\right] - \frac{\sigma_\delta^2}{I_{ch}}$$

Adding  $\sigma_\delta^2$  to both sides of the equality and arranging the terms, we get Eq(2).

## B Tables and Figures

Table 1: GLS regression results for standardized height in Coastal ecozone.

| <i>Variable</i>  | <i>Coef.</i> | <i>SE</i> |
|--|--------------|-----------|
| Intercept  | 77.48        | 0.73      |
| (Max years of educ for HH)*(Head has no educ)                      | 1.06         | 0.16      |
| (Head's years of educ)*(Water from dug well)                       | 0.50         | 0.11      |
| (# professional women in HH)*(Head has some secondary educ)        | -16.41       | 2.85      |
| (# women in HH for less than 5 years)*(# members aged 0)           | -8.66        | 1.53      |
| (Head's years of educ)*(Total deaths of children in HH last 12mth) | -0.67        | 0.10      |
| (Child deaths over live births)*(Age of spouse)                    | 0.35         | 0.07      |
| (Rain water)*(Head completed primary educ)                         | 10.83        | 1.68      |
| (Sex=Female)*(Head has some secondary educ)                        | 5.95         | 1.01      |
| (Roof made of rock)*(Age=4)  | -8.91        | 0.91      |
| (Rain water)*(Age=3)   | -6.44        | 1.21      |
| (Water not from pipe/well/bottle)*(Age=0)                          | 11.95        | 2.32      |
| (Ratio of female in HH)*(Age=3)                                    | -0.16        | 0.03      |
| (# women in HH worked last 12mth)*(Variance of NDVI in Nov)        | 12.41        | 3.15      |
| (Toilet on premise)*(Distance to river)                            | 2.97E-04     | 6.18E-05  |
| Ratio of head in village with secondary education                  | 21.98        | 4.45      |
| (Age of head)*(Water from dug well)                                | -0.09        | 0.02      |
| (# women worked)*(Child death over births)                         | -22.71       | 4.94      |
| (No women in HH have educ)*(Head has some secondary educ)          | -5.67        | 1.91      |
| (Head is male)*(age=0)   | 3.05         | 0.67      |

Table 2: GLS regression results for standardized weight in Coastal ecozone.

| <i>Variable</i>   | <i>Coef.</i> | <i>SE</i> |
|---|--------------|-----------|
| Intercept   | 10.21        | 0.21      |
| (Max years of educ for HH)*(Head has no educ)                               | 0.20         | 0.04      |
| (Head's years of educ)*(Water from dug well)                                | 0.11         | 0.03      |
| (# women in HH for less than 5 years)*(# members aged 0)                    | -2.07        | 0.38      |
| (Ratio of those ever attended school)*(Max year of educ for females in HH)  | 1.74E-03     | 4.32E-04  |
| (Roof made of rock)*(Age=4)   | -1.34        | 0.25      |
| (# members aged 0-4)*(Age=0)  | 0.85         | 0.10      |
| (Floor material wood)*(Area of rice field in village)                       | 3.99E-04     | 7.21E-05  |
| (Max female educ in HH is some primary educ)*(Δ ag. land b/w 1993 and 1997) | -7.13E-08    | 1.72E-08  |
| (Flood prone commune)*(Age=4)   | 1.88         | 0.34      |
| (# members aged 0-4)*(Water from dug well)                                  | -0.43        | 0.10      |
| (Age of head)*(toilet not on premise)                                       | -0.02        | 3.78E-03  |
| (# younger children)*(Roof made of wood/plastic)                            | 1.15         | 0.23      |
| (# students)*(Child death over live births)                                 | -1.20        | 0.35      |
| (toilet not on premise)*(head is male)                                      | 0.63         | 0.16      |
| (Roof made of wood/plastic)*(Age=3)   | -0.95        | 0.25      |
| (Head has some primary educ)*(Age=4)  | -0.96        | 0.23      |
| (Head is male)*(Age=1)  | -0.73        | 0.22      |
| (# members 65 or over)*(Age=0)  | 1.95         | 0.48      |
| (Max year educ. for females)*(Age=3)  | -0.21        | 0.04      |

Table 3: Comparison of the ecozone-level estimates. Standard error for CDHS Only was calculated by 100-time two-stage bootstrapping.

| Indicator   | Ecozone  | CDHS Only |      | This study |      |
|-------------|----------|-----------|------|------------|------|
|             |          | Mean      | SE   | Mean       | SE   |
| Stunting    | Urban    | 37.89     | 3.30 | 40.90      | 1.74 |
|             | Plain    | 47.58     | 2.49 | 50.62      | 1.92 |
|             | Tonlesap | 43.23     | 2.09 | 44.69      | 1.75 |
|             | Coastal  | 47.21     | 5.52 | 49.65      | 2.25 |
|             | Plateau  | 47.10     | 2.99 | 47.26      | 1.64 |
| Underweight | Urban    | 39.58     | 2.93 | 39.66      | 1.93 |
|             | Plain    | 47.80     | 2.44 | 46.35      | 1.73 |
|             | Tonlesap | 45.84     | 1.95 | 44.08      | 2.50 |
|             | Coastal  | 38.95     | 5.28 | 38.70      | 2.16 |
|             | Plateau  | 46.37     | 3.87 | 46.24      | 1.77 |

Table 4: Commune-level regression of malnutrition indicators on logarithmic mean consumption and inequality. Regressions were run separately. N=1594.

| Variable       | Without Provincial Dummies |         |             |         | With Provincial Dummies |         |             |         |
|----------------|----------------------------|---------|-------------|---------|-------------------------|---------|-------------|---------|
|                | Stunting                   |         | Underweight |         | Stunting                |         | Underweight |         |
|                | Coef                       | T-value | Coef        | T-value | Coef                    | T-value | Coef        | T-value |
| Intercept      | 0.528                      | 30.69   | 0.505       | 32.39   | 0.455                   | 22.47   | 0.507       | 27.09   |
| $\log \bar{Y}$ | -0.107                     | -4.50   | -0.082      | -3.78   | -0.002                  | -0.82   | -0.003      | -1.41   |
| GE(0)          | -0.005                     | -2.01   | -0.007      | -3.22   | -0.054                  | -2.40   | -0.061      | -2.90   |
| $R^2$          | 0.02                       |         | 0.03        |         | 0.21                    |         | 0.18        |         |

Table 5: Commune-level regression of average severity of malnutrition  $\frac{P^2}{P^0}$  on logarithmic mean consumption and inequality. Right two columns also includes  $P^0$ . Regressions were run separately. N=1594.

| Variable       | Stunting |         | Underweight |         | Stunting  |         | Underweight |         |
|----------------|----------|---------|-------------|---------|-----------|---------|-------------|---------|
|                | Coef     | T-value | Coef        | T-value | Coef      | T-value | Coef        | T-value |
| Intercept      | 0.0074   | 21.31   | 0.0294      | 21.87   | 0.0032    | 9.46    | 0.0053      | 4.30    |
| $\log \bar{Y}$ | 0.0001   | 0.35    | -0.0002     | -0.16   | -4.46E-05 | -1.26   | -2.77E-05   | -0.23   |
| GE(0)          | -0.0001  | -1.50   | -0.0002     | -1.10   | 0.0006    | 1.96    | 0.0026      | 2.34    |
| $P^0$          |          |         |             |         | 0.0093    | 25.51   | 0.0476      | 35.08   |
| $R^2$          | 0.19     |         | 0.26        |         | 0.43      |         | 0.59        |         |

Table 6: Number of communes by poverty rate (row) and prevalence of underweight (column).

|           | 0.00-0.30 | 0.30-0.45 | 0.45-0.50 | 0.50-0.55 | 0.55-1.00 | Total |
|-----------|-----------|-----------|-----------|-----------|-----------|-------|
| 0.00-0.30 | 43        | 169       | 235       | 52        | 50        | 549   |
| 0.30-0.50 | 19        | 126       | 295       | 62        | 53        | 555   |
| 0.50-1.00 | 23        | 122       | 208       | 81        | 56        | 490   |
| Total     | 85        | 417       | 738       | 195       | 159       | 1594  |

Table 7: Decomposition of health and consumption inequality as measured by GE(0). Health inequality is corrected for the natural inequality as proposed by Pradhan et al. (2003). Unit record is individual for standardized height and weight, and household for consumption. Standard errors are in parenthesis. The number of observations for unit record is 1424907 for standardized height and weight and 213054 for consumption. All figures are in percentage.

|                        | Cambodia    | Province      | District      | Commune       | Village      | Individual/Household |
|------------------------|-------------|---------------|---------------|---------------|--------------|----------------------|
| Standardized Height    | 0.00 (0.00) | 1.29 (0.41)   | 2.94 (0.48)   | 7.69 (1.04)   | 10.87 (1.13) | 100.00 (0.00)        |
| Standardized Weight    | 0.00 (0.00) | 1.49 (0.40)   | 3.65 (0.54)   | 8.10 (1.05)   | 12.84 (1.29) | 100.00 (0.00)        |
| Consumption            | 0.00 (0.00) | 18.69 (16.32) | 28.84 (15.84) | 43.19 (13.16) | 64.14 (8.68) | 100.00 (0.00)        |
| Number of Observations | 1           | 24            | 180           | 1594          | 13320        | 1424907/2130544      |

Table 8: Decomposition of the variance of FGT measures for parameter 0, 1 and 2. Unit record for standardized height and weight is individual and household for consumption. Standard errors are in parenthesis. The number of observations for unit record is 1424907 for standardized height and weight and 213054 for consumption. All figures are in percentage.

| Indicator              | Parameter | Cambodia    | Province    | District     | Commune      | Village      | Individual/Household |
|------------------------|-----------|-------------|-------------|--------------|--------------|--------------|----------------------|
| Standardized Height    | 0         | 0.00 (0.00) | 0.95 (0.29) | 1.92 (0.31)  | 4.23 (0.45)  | 6.28 (0.50)  | 100.00 (0.00)        |
|                        | 1         | 0.00 (0.00) | 0.95 (0.31) | 2.10 (0.38)  | 4.98 (0.70)  | 7.22 (0.75)  | 100.00 (0.00)        |
|                        | 2         | 0.00 (0.00) | 0.73 (0.24) | 1.70 (0.32)  | 4.31 (0.75)  | 6.40 (0.81)  | 100.00 (0.00)        |
| Standardized Weight    | 0         | 0.00 (0.00) | 0.75 (0.20) | 1.85 (0.27)  | 3.92 (0.47)  | 6.48 (0.57)  | 100.00 (0.00)        |
|                        | 1         | 0.00 (0.00) | 0.78 (0.24) | 2.04 (0.34)  | 4.31 (0.70)  | 7.01 (0.81)  | 100.00 (0.00)        |
|                        | 2         | 0.00 (0.00) | 0.59 (0.21) | 1.52 (0.30)  | 3.20 (0.62)  | 5.49 (0.71)  | 100.00 (0.00)        |
| Consumption            | 0         | 0.00 (0.00) | 7.10 (0.50) | 12.00 (0.71) | 19.06 (0.88) | 37.14 (1.16) | 100.00 (0.00)        |
|                        | 1         | 0.00 (0.00) | 8.20 (0.86) | 14.29 (1.36) | 23.15 (1.76) | 46.72 (2.01) | 100.00 (0.00)        |
|                        | 2         | 0.00 (0.00) | 7.20 (1.00) | 12.98 (1.81) | 21.53 (2.36) | 45.87 (2.82) | 100.00 (0.00)        |
| Number of Observations |           | 1           | 24          | 180          | 1594         | 13320        | 1424907/2130544      |

Table 9: Decomposition of Gini for standardized height into within-location, between-location and residual components. All figures are in percentage.

|                    | Cambodia | Province | District | Commune | Village | Individual |
|--------------------|----------|----------|----------|---------|---------|------------|
| $\frac{NCI_W}{GI}$ | 100.00   | 3.59     | 0.42     | 0.04    | 0.01    | 0.00       |
| $\frac{NCI_B}{GI}$ | 0.00     | 5.62     | 8.20     | 11.94   | 14.73   | 100.00     |
| $\frac{CI_R}{GI}$  | 0.00     | 90.79    | 91.38    | 88.02   | 85.26   | 0.00       |



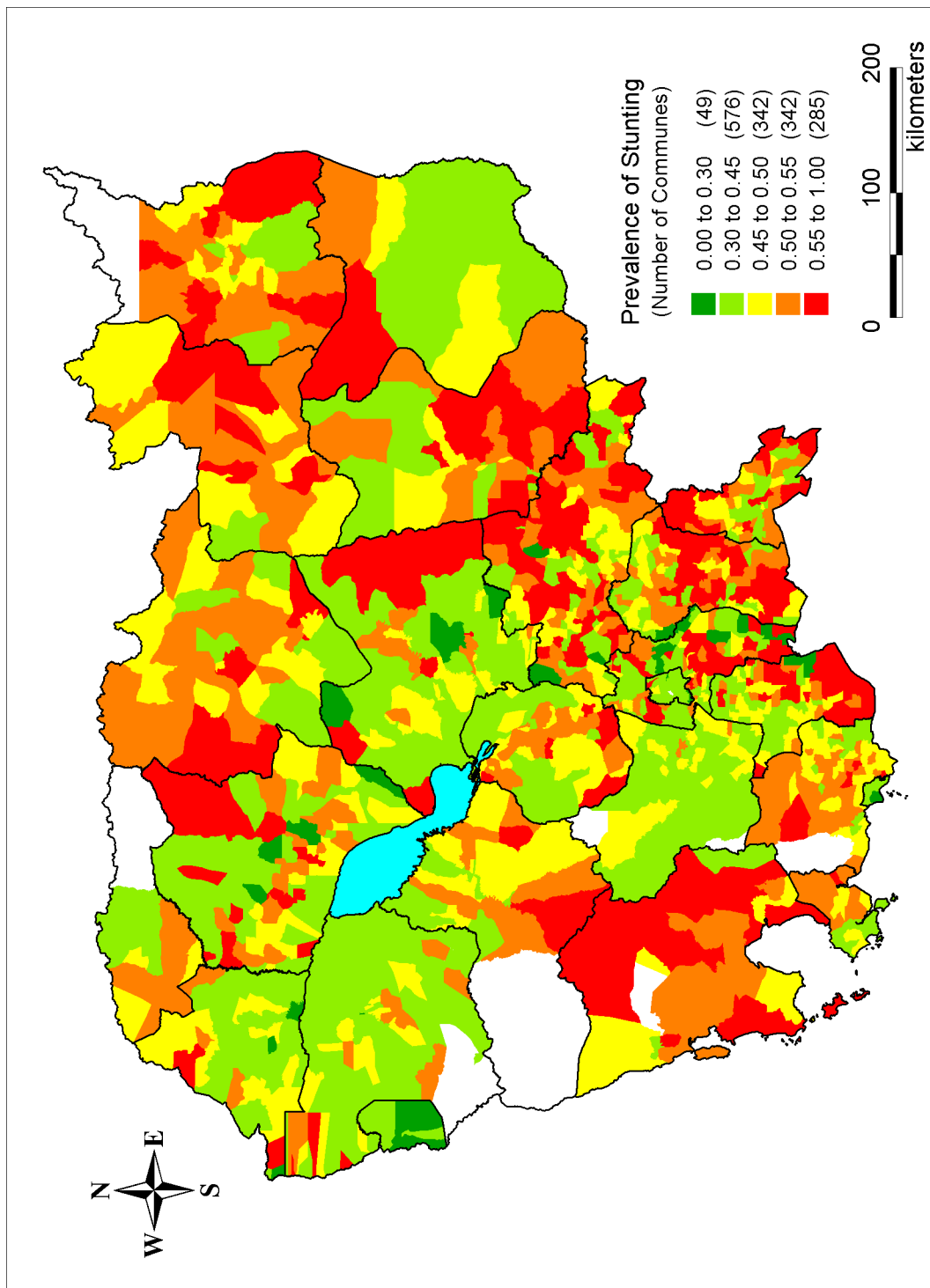


Figure 1: Commune-level prevalence of stunting for children under five in Cambodia.

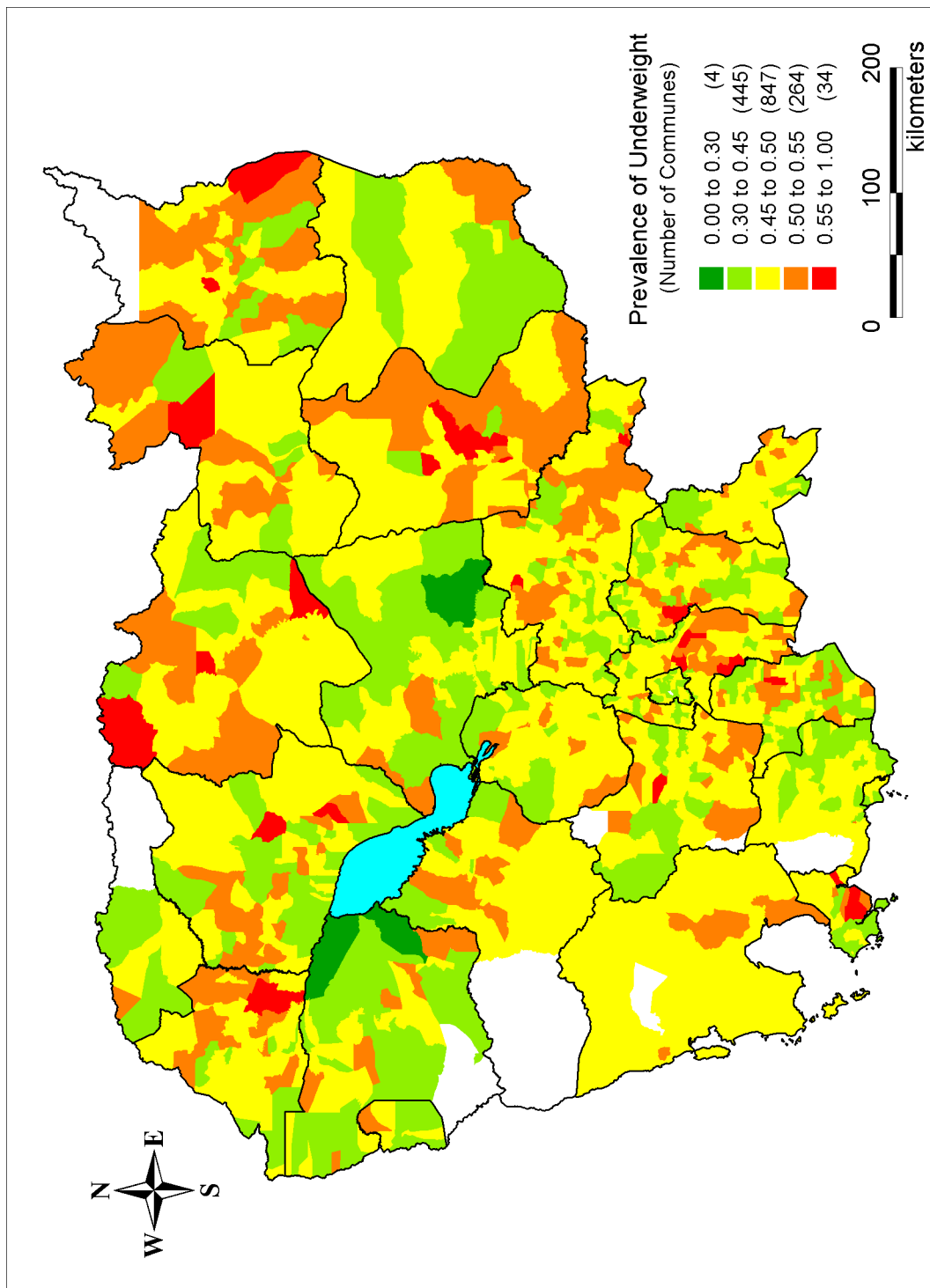


Figure 2: Commune-level prevalence of underweight for children under five in Cambodia.

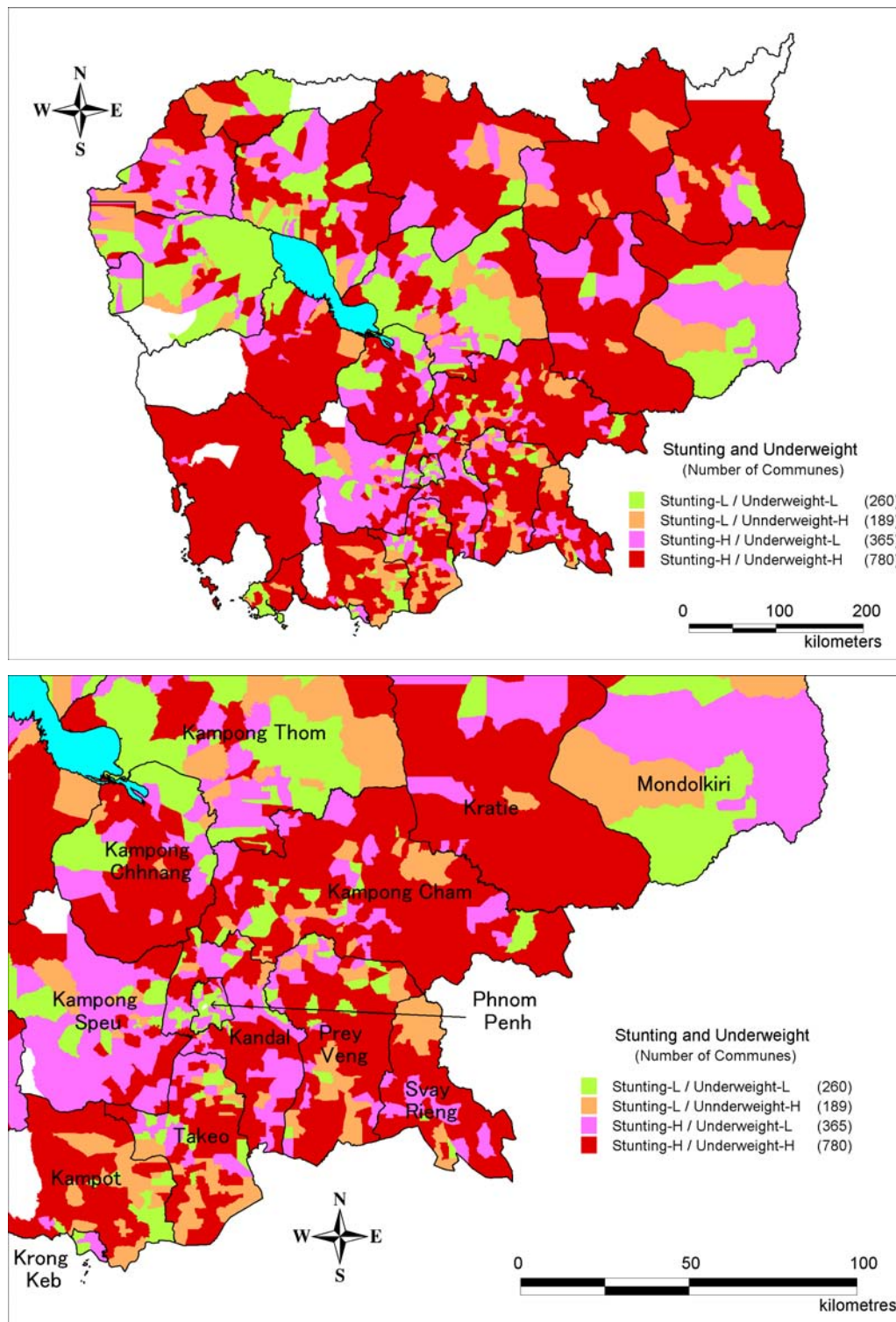


Figure 3: (a) Prevalence of stunting and underweight in Cambodia (top), and (b) near Phnom Penh (bottom). H and L respectively mean above and below 45 percent.

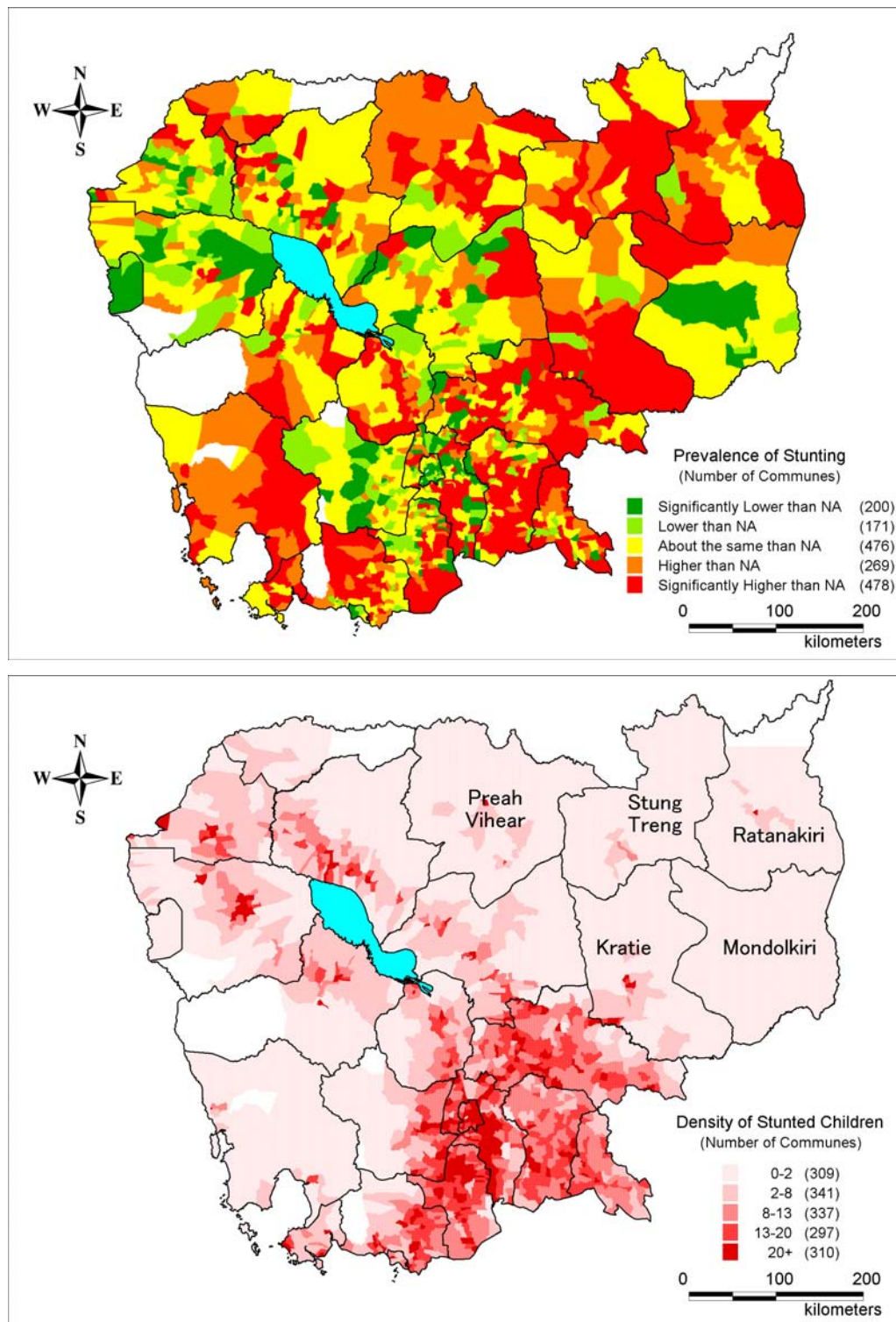


Figure 4: (a) Prevalence of stunting in comparison with the national average (NA) (top), and (b) the number of stunted children per square kilometers (bottom).



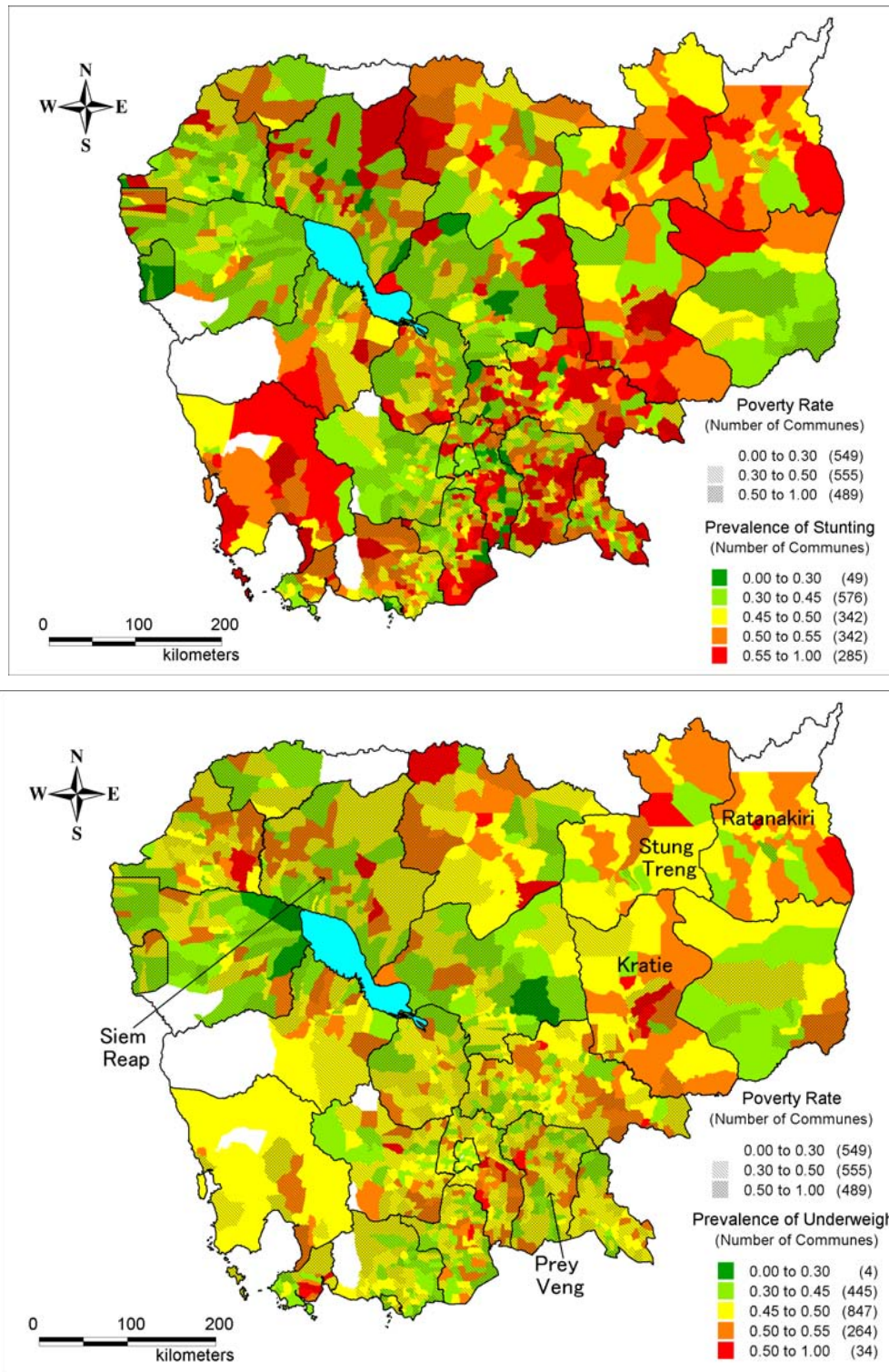


Figure 5: (a) Poverty rate and prevalence of stunting (top), and (b) poverty rate and prevalence of underweight in Cambodia.

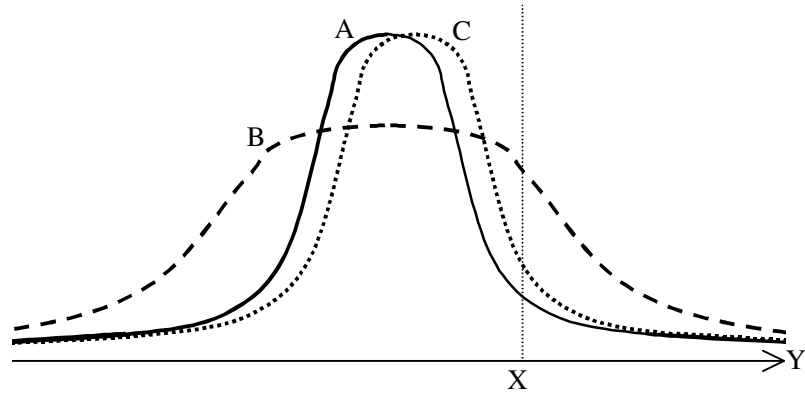


Figure 6: An illustration of the relationship between consumption and malnutrition. Horizontal axis measures consumption and vertical axis measures probability density.  $X$  is the consumption threshold below which a child is malnourished.

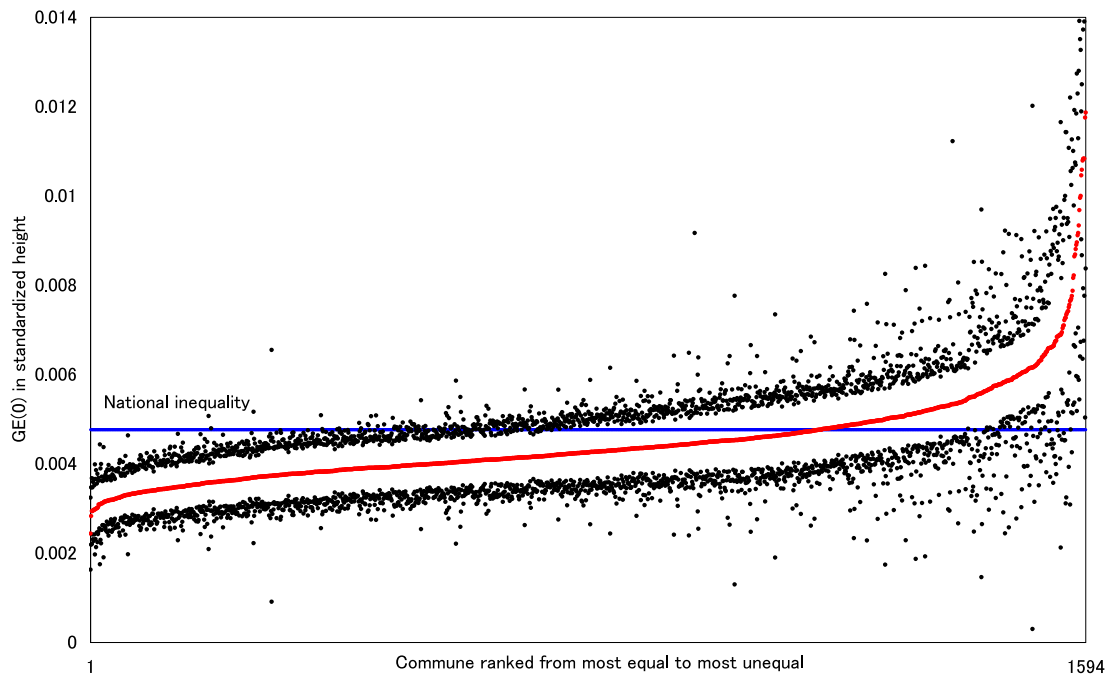


Figure 7: Distribution of commune-level  $GE(0)$  in standardized height.

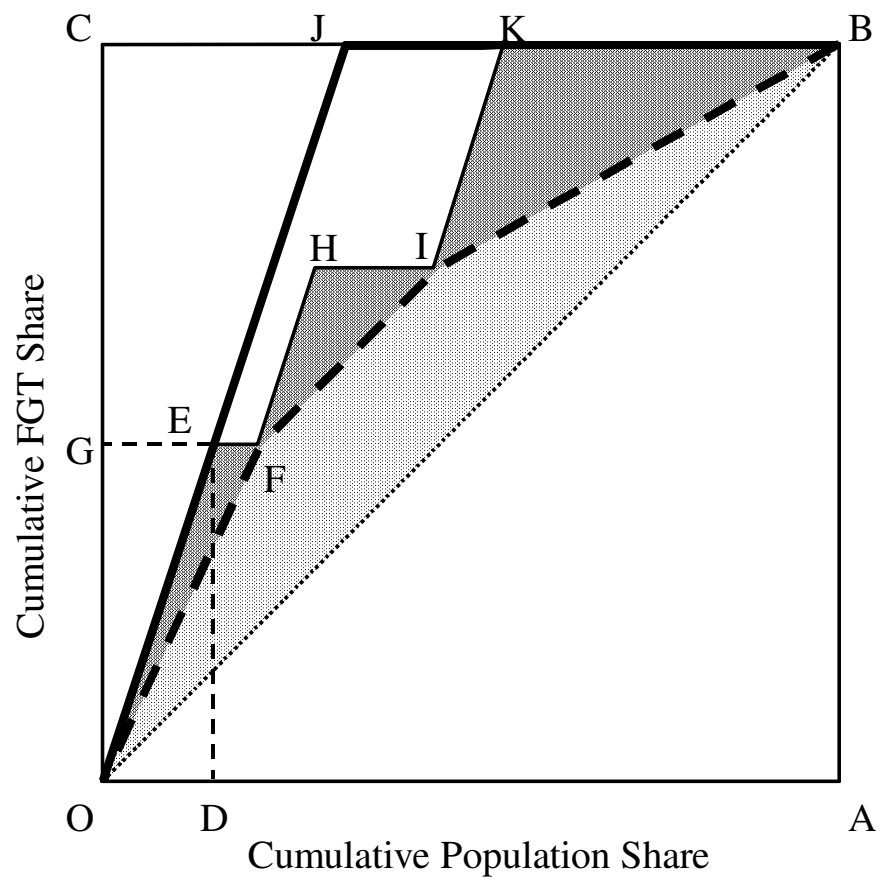


Figure 8: An example of concentration curve for  $P^0$ .

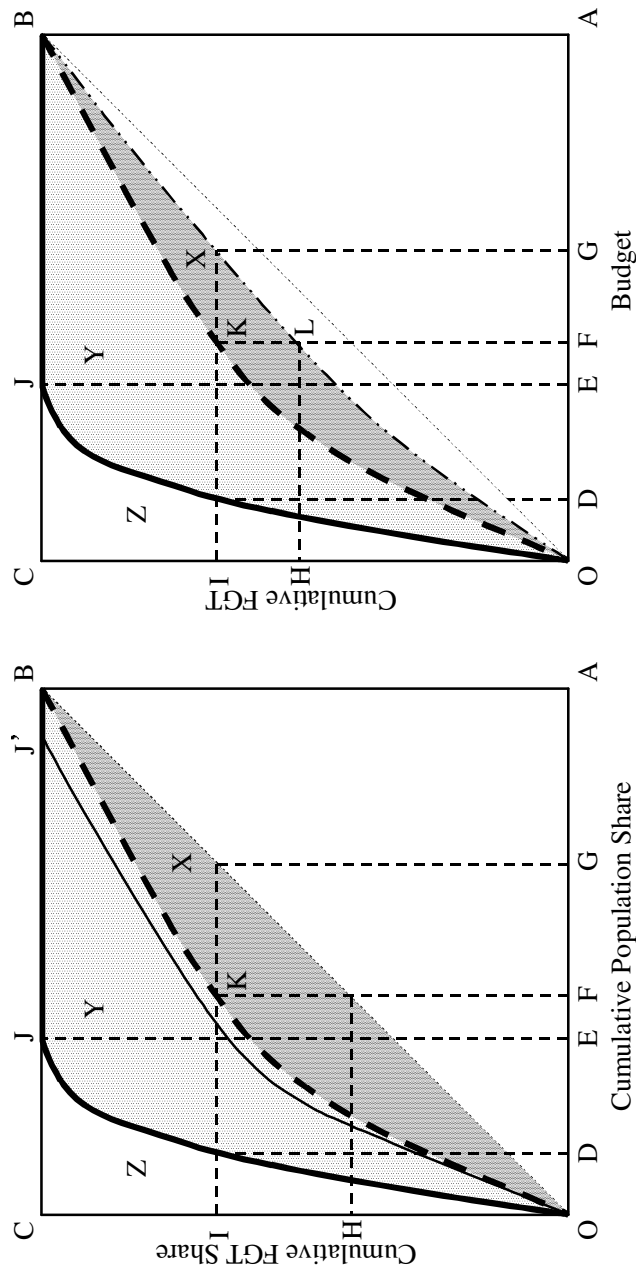


Figure 9: (a) An example of concentration curve (left) and (b) an example of adjusted concentration curve (right).



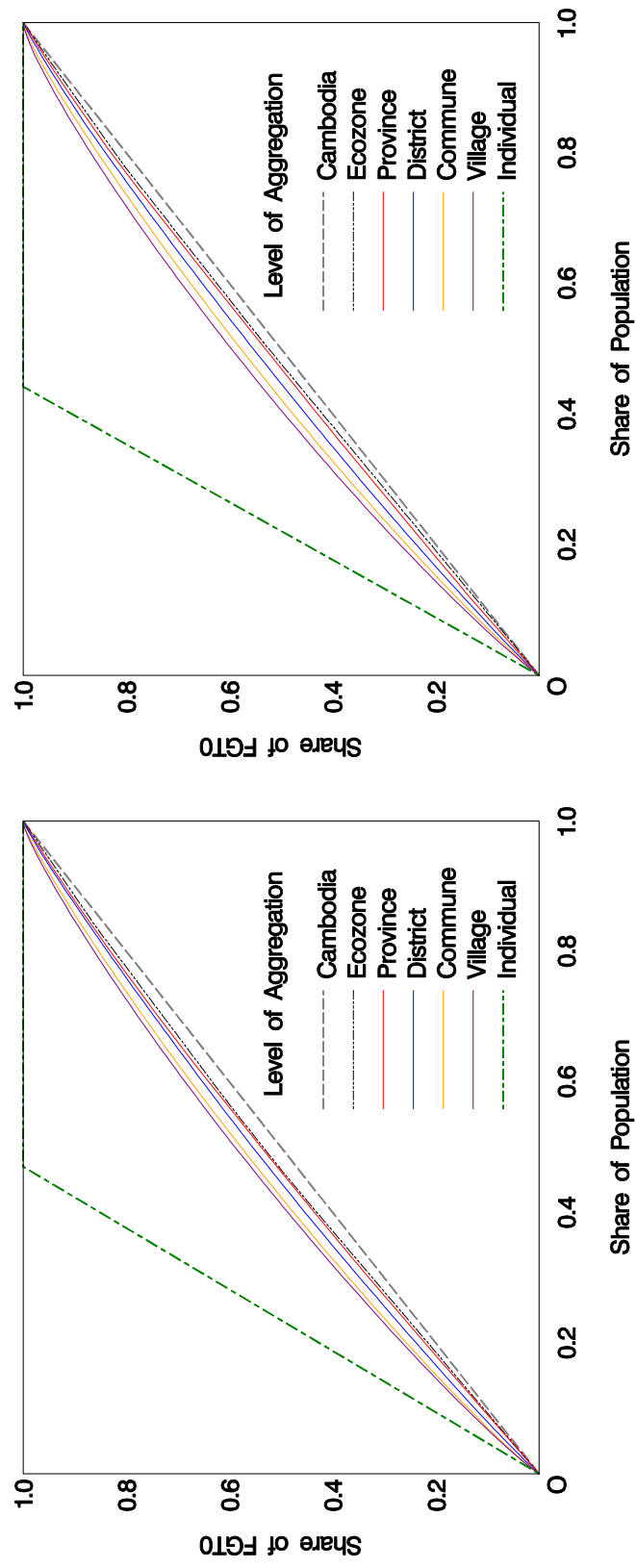


Figure 10: (a) Concentration curve for stunting (left), and (b) concentration curve for underweight (right). The horizontal axis measures the cumulative share of population and the vertical axis measures the cumulative share of malnourished children.

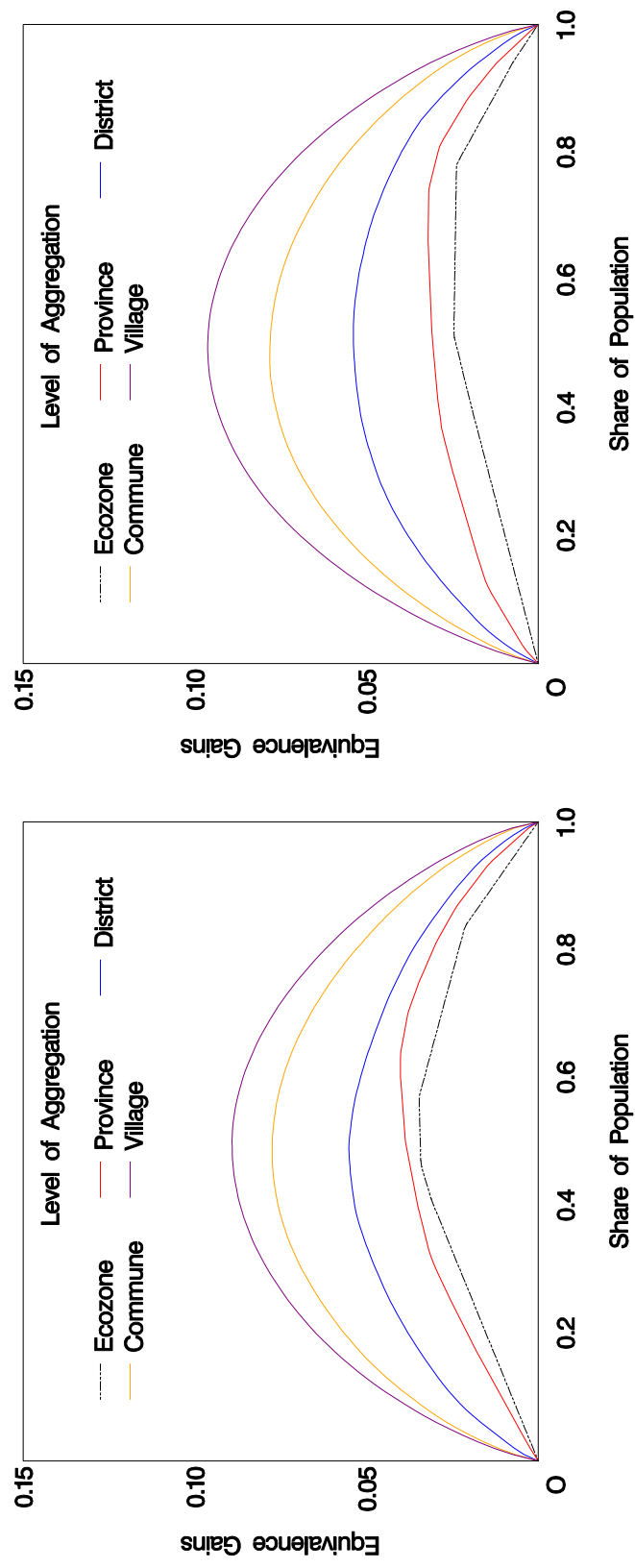


Figure 11: (a) Equivalence gains for stunting (left), and (b) equivalence gains for underweight (right). The horizontal axis measures the share of population to be targeted and the vertical axis measures equivalence gains.

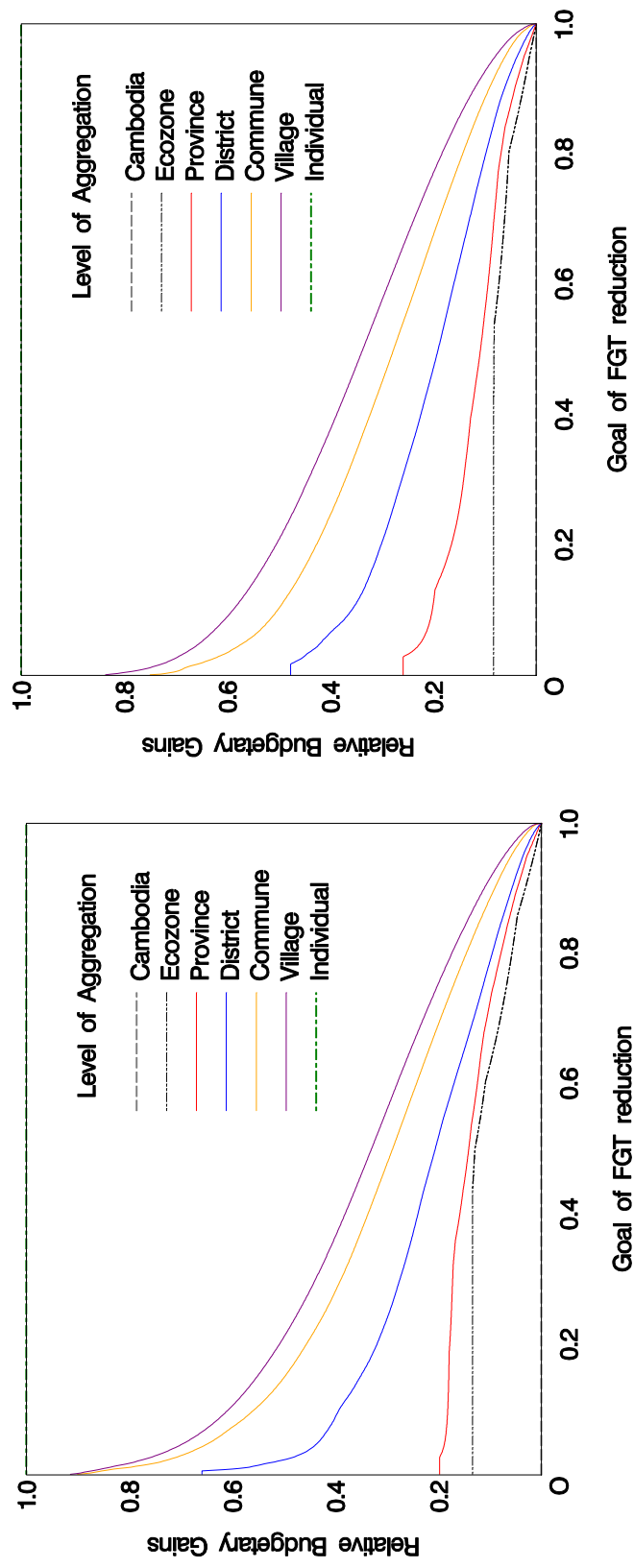


Figure 12: (a) Relative budgetary gains for stunting (left), and (b) relative budgetary gains for underweight (right). The horizontal axis measures the goal of FGT reduction and the vertical axis measures the relative budgetary gains.

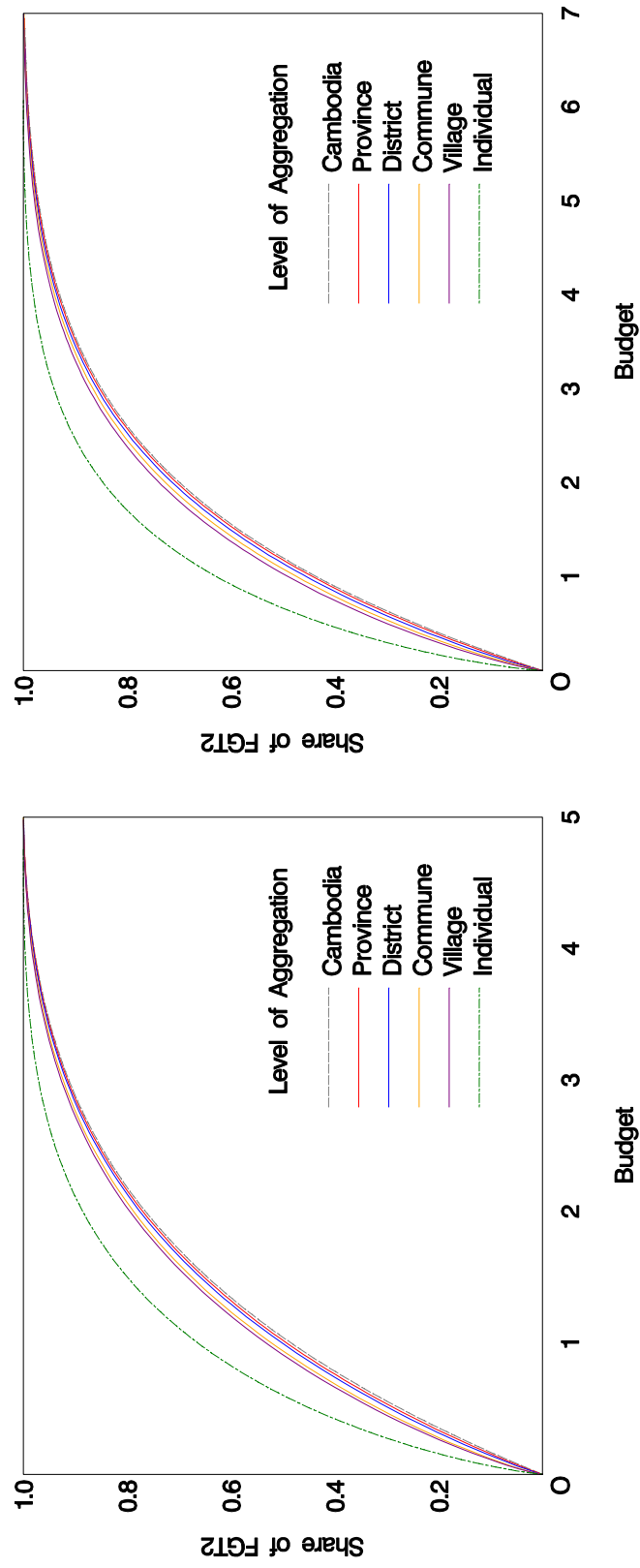


Figure 13: (a) Adjusted concentration curve for stunting (left) and (b) adjusted concentration curve for underweight. The horizontal axis measures the budget and the vertical axis measures the reduction in FGT.

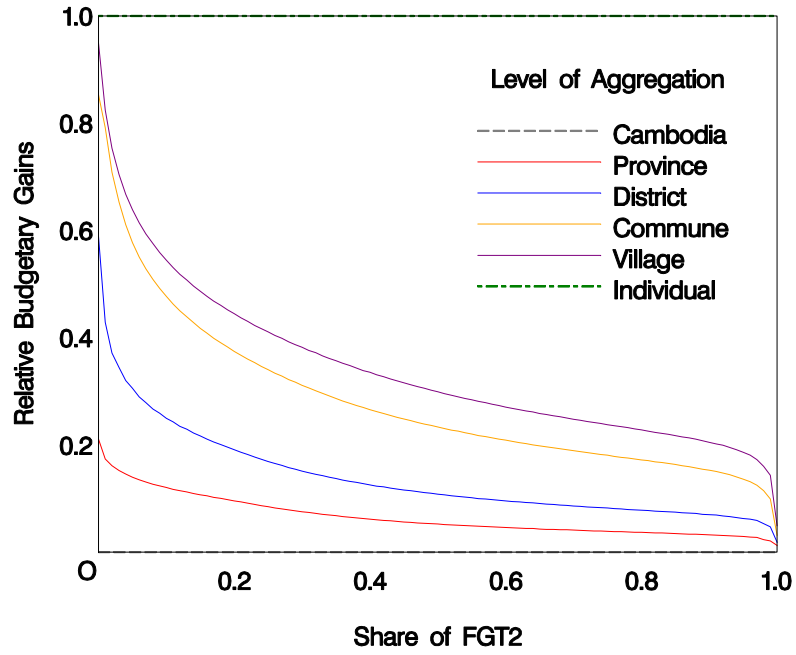


Figure 14: Relative budgetary gains for adjusted concentration curve in Figure 13. The horizontal axis measures the goal of reduction in FGT measure  $P^{2,(1)}$ . The vertical axis measures relative budgetary gains.